



# An Infrared Search for Kilonovae with the WINTER Telescope. I. Binary Neutron Star Mergers

Danielle Frostig<sup>1</sup>, Sylvia Biscoveanu<sup>1,2</sup>, Geoffrey Mo<sup>1,2</sup>, Viraj Karambelkar<sup>3</sup>, Tito Dal Canton<sup>4</sup>, Hsin-Yu Chen<sup>1,2</sup>, Mansi Kasliwal<sup>3</sup>, Erik Katsavounidis<sup>1,2</sup>, Nathan P. Lourie<sup>1</sup>, Robert A. Simcoe<sup>1</sup>, and Salvatore Vitale<sup>1,2</sup>

<sup>1</sup> Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA; [frostig@mit.edu](mailto:frostig@mit.edu), [sbisco@mit.edu](mailto:sbisco@mit.edu)

<sup>2</sup> LIGO Laboratory, Massachusetts Institute of Technology, 185 Albany St, Cambridge, MA 02139, USA

<sup>3</sup> Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA

<sup>4</sup> Université Paris-Saclay, CNRS/IN2P3, IJCLab, F-91405 Orsay, France

Received 2021 September 22; revised 2021 November 17; accepted 2021 December 14; published 2022 February 21

## Abstract

The Wide-Field Infrared Transient Explorer (WINTER) is a new 1 deg<sup>2</sup> seeing-limited time-domain survey instrument designed for dedicated near-infrared follow-up of kilonovae from binary neutron star (BNS) and neutron star–black hole mergers. WINTER will observe in the near-infrared *Y*, *J*, and short-*H* bands (0.9–1.7  $\mu$ m, to  $J_{AB} = 21$  mag) on a dedicated 1 m telescope at Palomar Observatory. To date, most prompt kilonova follow-up has been in optical wavelengths; however, near-infrared emission fades more slowly and depends less on geometry and viewing angle than optical emission. We present an end-to-end simulation of a follow-up campaign during the fourth observing run (O4) of the LIGO, Virgo, and KAGRA interferometers, including simulating 625 BNS mergers, their detection in gravitational waves, low-latency and full parameter estimation skymaps, and a suite of kilonova lightcurves from two different model grids. We predict up to five new kilonovae independently discovered by WINTER during O4, given a realistic BNS merger rate. Using a larger grid of kilonova parameters, we find that kilonova emission is  $\approx 2$  times longer lived and red kilonovae are detected  $\approx 1.5$  times further in the infrared than in the optical. For 90% localization areas smaller than 150 (450) deg<sup>2</sup>, WINTER will be sensitive to more than 10% of the kilonova model grid out to 350 (200) Mpc. We develop a generalized toolkit to create an optimal BNS follow-up strategy with any electromagnetic telescope and present WINTER’s observing strategy with this framework. This toolkit, all simulated gravitational-wave events, and skymaps are made available for use by the community.

*Unified Astronomy Thesaurus concepts:* Infrared telescopes (794); Gravitational wave astronomy (675); Neutron stars (1108)

## 1. Introduction

During the second Advanced LIGO–Virgo observing run (O2; Aasi et al. 2015; Acernese et al. 2015), the binary neutron star (BNS) merger GW170817 led to the first coincident detection of gravitational waves (GWs) with electromagnetic (EM) radiation spanning X-ray to radio frequencies (Abbott et al. 2017a, 2017b). GW170817 marked the first direct evidence of a kilonova—a thermal transient powered by rapid neutron capture (*r*-process) nucleosynthesis in the neutron-rich ejecta of BNS mergers (Li & Paczynski 1998; Rosswog 2005; Metzger et al. 2010; Roberts et al. 2011; Barnes & Kasen 2013; Kasen et al. 2013; Tanaka & Hotokezaka 2013; Grossman et al. 2014; Barnes et al. 2016; Abbott et al. 2017c; Coulter et al. 2017; Evans et al. 2017; Goldstein et al. 2017; Haggard et al. 2017; Hallinan et al. 2017; Margutti et al. 2017; Troja et al. 2017; Kasliwal et al. 2022; Metzger 2020). Kilonovae offer a unique laboratory in which to study the production of heavy elements via the *r*-process, probe the neutron star equation of state (Margalit & Metzger 2017; Coughlin et al. 2018a, 2019a; Breschi et al. 2021), and provide a new class of standard sirens for resolving the Hubble tension (Holz & Hughes 2005; Abbott et al. 2017d; Chen et al. 2018; Coughlin et al. 2020a;

Dietrich et al. 2020). The combination of GW alerts and electromagnetic follow-up provides a new tool to expand the small data set of known kilonovae from BNS (although see Tanvir et al. 2013; Troja et al. 2018; Fong et al. 2021) and neutron star–black hole (NSBH) mergers (Abbott et al. 2021a).

Despite the successful observation of GW170817 in O2, no new, confirmed electromagnetic counterparts to BNS or NSBH mergers were discovered during the third Advanced LIGO–Virgo observing run (O3; Abbott et al. 2020c, 2021a, 2021b; Coughlin et al. 2020b; Kasliwal et al. 2020; Zhu et al. 2021a). Many ultraviolet/visible/infrared (UVIOR) teams undertook follow-up observations of the O3 LIGO alerts, including the ASAS-SN (Kochanek et al. 2017), ATLAS (Tonry et al. 2018), BOOTES (Hu et al. 2021), DDOTI (Becerra et al. 2021), DES-GW (Soares-Santos et al. 2017), ENGRAVE (Levan 2020), GOTO (Gompertz et al. 2020), GRANDMA (Antier et al. 2020), GROWTH (Kasliwal et al. 2020), MASTER-Net (Lipunov et al. 2017), SAGUARO (Paterson et al. 2021), Swift UVOT (Oates et al. 2021), and VINROUGE (Ackley et al. 2020) teams. In addition to searches for UVOIR kilonovae, the Fermi and Swift (Page et al. 2020) spacecraft conducted gamma-ray and X-ray follow-up, respectively, and the Australian Square Kilometre Array Pathfinder searched in radio wavelengths (Dobie et al. 2019). During O3, spurious GW alerts and alert retractions complicated electromagnetic follow-up strategies (Coughlin et al. 2020c). Additionally, models predict many of the BNS mergers detected in GWs in



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

O3 produced kilonovae that were too faint to be detected by the UVOIR facilities conducting follow-up at the time (Kasliwal et al. 2020; Zhu et al. 2021a).

Optical-wavelength searches dominated the UVOIR O3 GW follow-up landscape, with dozens of optical telescopes across the globe and in orbit triggered during O3 for GW follow-up. Theoretical models and observational constraints from O3 nondetections predict that the blue, optical emission from a kilonova is angle-dependent, fades rapidly ( $<1$  week), and may not be present in all BNS or NSBH mergers (Kasen et al. 2013, 2017; Barnes et al. 2016; Kasliwal et al. 2020; Metzger 2020). By contrast, the near-infrared emission is expected to be isotropic, long-lived ( $>1$  week), and ubiquitous in models regardless of mass ratio, viewing angle, or remnant lifetime (Kasen et al. 2017). Models predict the detection rates of kilonovae in the near-infrared could be up to  $\sim 8$ – $10$  times higher than in optical wave bands (Zhu et al. 2021b).

However, due to the high cost-per-pixel of detectors and bright sky backgrounds, the dynamic infrared sky remains largely underexplored compared to optical wavelengths. Existing time-domain, infrared surveys are either restricted to small areas on sky (e.g., the VISTA Variables in the Via Lactea survey covering  $520 \text{ deg}^2$  (Catelan et al. 2011, p. 145) or the UKIRT Deep Extragalactic Survey covering  $35 \text{ deg}^2$  (Lawrence et al. 2007)) or are relatively shallow (e.g., Palomar Gattini IR,  $J \sim 16$  mag; De et al. 2020). The lack of deep, all-sky reference images limits the number of GW events that can be followed up in the infrared.

The Wide-field Infrared Transient Explorer (WINTER) is a new instrument that will perform the first near-infrared, all-sky survey to  $J_{AB} = 21$  mag and is specially built for GW follow-up (Frostig et al. 2020; Lourie et al. 2020). WINTER will operate on a dedicated 1 m telescope at Palomar Observatory that was commissioned in 2021 June with an optical camera on one port, with the infrared WINTER instrument to be added on the second port in late 2021. WINTER’s wide,  $1 \text{ deg}^2$  field of view can quickly tile the median expected fourth observing run (O4)  $33 \text{ deg}^2$  BNS localization contour (Abbott et al. 2020a) with rapid-response robotic observing. With three near-infrared filters in the  $Y$ ,  $J$ , and shortened- $H$  bands (centered at 1.0, 1.2, and  $1.6 \mu\text{m}$ , respectively), WINTER is designed to discover kilonovae and observe them for two weeks or more (Frostig et al. 2020).

As an alternative to the traditional, but expensive, mercury–cadmium–telluride (HgCdTe) detectors that dominate the near-infrared landscape, WINTER employs cheaper indium–gallium–arsenide (InGaAs) detectors new to astronomical instrumentation. A prototype instrument confirmed InGaAs detectors achieve background-limited near-infrared photometry without cryogenic cooling (Simcoe et al. 2019), which is required for HgCdTe detectors. WINTER will head a new class of InGaAs-based near-infrared astronomical instruments coming online in the next decade, including the DREAMS telescope in the Southern Hemisphere (Travouillon et al. 2020). The upcoming HgCdTe-based PRIME-focus Infrared Microlensing Experiment telescope will also join WINTER and DREAMS in an upcoming effort to deepen our understanding of the near-infrared time-domain sky.

In this paper, we present an end-to-end simulation of WINTER’s performance in O4 and make a case for infrared follow-up of BNS GW signals. Existing simulations such as Abbott et al. (2018) also present predictions for BNS merger

rates detected by the global GW network in O4, but do not model any electromagnetic counterparts. Petrov et al. (2022) repeats the analysis in Abbott et al. (2018) with more realistic parameters and studies the resulting BNS lightcurves using only optical-wavelength kilonova models. Our study additionally takes into account uncertainty introduced by the use of realistic GW matched-filter pipelines, compares any potential differences between the use of low- and medium-latency skymaps, and is primarily focused on the infrared. We include only BNS mergers in this work and leave the study of NSBH mergers, which are expected to be especially promising to follow up in the infrared (Fernández et al. 2017; Zhu et al. 2020; Anand et al. 2021), for future work.

In Section 2 we describe the simulation, including modeling a population of BNS mergers, their respective skymaps, and WINTER follow-up observations. We present the results in Section 3. This leads to a discussion of the merits of studying kilonovae in the infrared and the planning of WINTER observations in Section 4, and we conclude in Section 5.

## 2. Methods

We simulate follow-up of BNS gravitational-wave signals with WINTER using the following procedure:

1. We generate a simulated BNS population and model a set of electromagnetic lightcurves from two different model grids (Kasen et al. 2017; Bulla 2019) for each merger event.
2. To model the gravitational-wave response to each event, we simulate Advanced LIGO (Aasi et al. 2015) and Virgo (Acernese et al. 2015) detector noise and network configurations expected for O4. We also include the Japanese interferometer KAGRA (Somiya 2012; Aso et al. 2013; Akutsu et al. 2018), which joined the global network of ground-based gravitational-wave detectors at the end of the third observing run.
3. We perform a low-latency search for gravitational-wave events in our simulated data using the PyCBC Live pipeline (Nitz et al. 2018; Dal Canton et al. 2021). Low-latency skymaps are generated using the BAYESTAR software (Singer & Price 2016) to obtain the localization uncertainty contour for each event recovered by PyCBC Live. Full source characterization is then performed using the bilby parameter estimation pipeline (Ashton et al. 2019; Romero-Shaw et al. 2020) to obtain medium-latency skymaps. Our simulated PyCBC Live results, BAYESTAR skymaps, and bilby posteriors are publicly available on Zenodo at Frostig et al. (2021).
4. We simulate WINTER follow-up observations searching the resultant BAYESTAR and bilby skymaps. If WINTER successfully observes the true location of the event, we then verify that the event is bright enough to qualify as a statistically significant detection.

### 2.1. Simulated BNS Population

The fourth observing run of the Advanced LIGO gravitational-wave interferometers (Aasi et al. 2015) is expected to have a duration of one year and include a four-detector network consisting of the two Advanced LIGO detectors (Aasi et al. 2015), Advanced Virgo (Acernese et al. 2015), and KAGRA (Somiya 2012; Aso et al. 2013; Abbott et al. 2018;

Akutsu et al. 2018). One major goal of this study is to determine how many BNS systems would be detected in gravitational waves and then successfully followed up with WINTER during O4. To this end, we simulate a realistic population of BNS systems that could be detected over the course of one year of observing with this detector network operating at its expected sensitivity during this observing run (Abbott et al. 2020b). A (quasi-circular) BNS system is fully described by 17 binary parameters,  $\theta$ , including the component masses, six-dimensional spins, and tidal deformabilities of the individual neutron stars, and the extrinsic parameters—the distance, inclination angle, R.A., decl., polarization angle, and time and phase at coalescence.

We simulate systems following the mass distribution providing the best fit to the population of Galactic double neutron stars and the components of GW170817 as determined in Farrow et al. (2019), which is parameterized in terms of the “slow” and “recycled” binary neutron star components. In the isolated channel for BNS formation (e.g., Tauris et al. 2017), the recycled neutron star forms first and gets spun up to a period of  $\sim 10$ – $100$  ms via accretion from its companion (Alpar et al. 1982; Radhakrishnan & Srinivasan 1982; Heuvel 2017). Conversely, the second-born neutron star does not experience this period of accretion, and spins down to a “slow” period of  $\mathcal{O}(1)$  s on a timescale of  $\sim 1$  Myr. The Galactic double pulsar PSR J0737-3039A/B serves as the canonical example for this type of system (Burgay et al. 2003; Lyne et al. 2004).

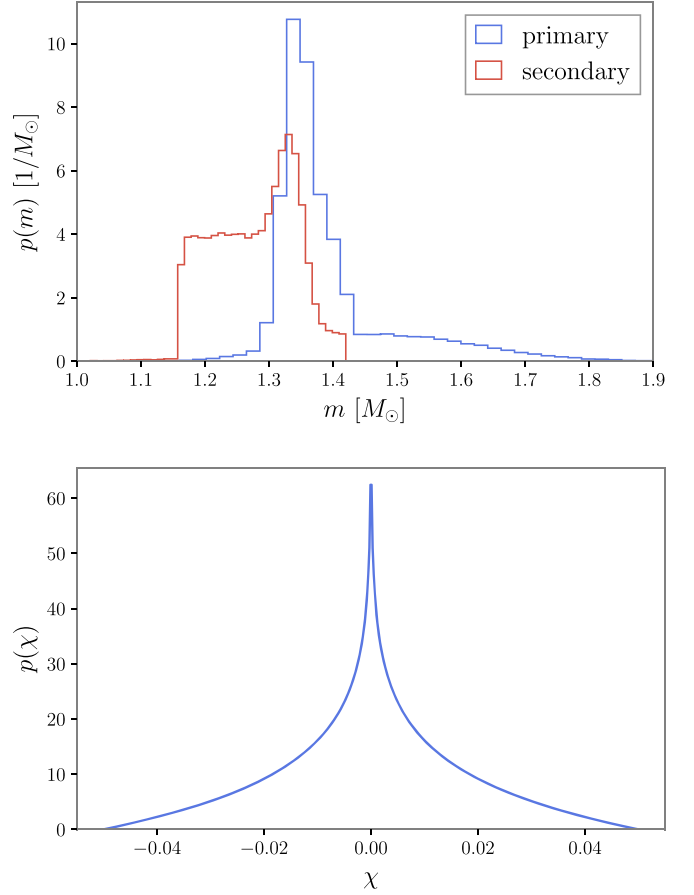
We model the mass distribution of the recycled neutron star,  $m_r$ , as a Gaussian mixture model, with mixing fraction  $\alpha = 0.68$  determining the fraction of systems in the first Gaussian:

$$p(m_s) = \alpha \mathcal{N}(m_r; \mu_1, \sigma_1) + (1 - \alpha) \mathcal{N}(m_r; \mu_2, \sigma_2). \quad (1)$$

The two Gaussians have means and widths given by  $\mu_1 = 1.34 M_\odot$ ,  $\sigma_1 = 0.02 M_\odot$ ,  $\mu_2 = 1.47 M_\odot$ , and  $\sigma_2 = 0.15 M_\odot$ . We draw the masses of the slow neutron star,  $m_s$ , in each binary from a uniform distribution between  $[1.16 M_\odot, 1.42 M_\odot]$  (Farrow et al. 2019). The primary and secondary masses are assigned via  $m_1 = \max(m_r, m_s)$ ,  $m_2 = \min(m_r, m_s)$ , the distributions for which are shown in the top panel of Figure 1. The tidal deformabilities of the neutron stars,  $\Lambda_i$ , are calculated from the masses assuming the AP4 equation of state (Akmal & Pandharipande 1997; Akmal et al. 1998).

We assume the dimensionless spins of the neutron stars  $\chi$  are aligned with the orbital angular momentum and drawn from the implied distribution on  $\chi_z$  assuming that the magnitudes are distributed uniformly on  $[0, 0.05]$  and the directions are isotropic. This spin prior choice is motivated by the maximum spin of observed Galactic double neutron star systems (Lorimer 2008). The resultant distribution is shown in the bottom panel of Figure 1. The inclination angle,  $\theta_{\text{JN}}$ , between the total angular momentum and the line of sight of the observer is drawn uniformly in  $\cos \theta_{\text{JN}}$ , and the orbital phase at coalescence and polarization angle also follow uniform distributions.

The detector-frame coalescence times,  $t_d$ , of the sources are distributed uniformly over the course of one year starting on 2022 January 1. We choose the sky locations and distances (and hence the redshifts) assuming the sources are distributed isotropically on the sky and uniformly in comoving volume and source-frame time, such that  $p(z, t_d) \propto \frac{dV_c}{dz} (1+z)^{-1}$  out to a luminosity distance of 400 Mpc, where  $V_c$  is the comoving



**Figure 1.** Top: probability density function for the distribution of primary and secondary source-frame BNS masses from Farrow et al. (2019). Bottom: probability density function for the distribution of dimensionless spins.

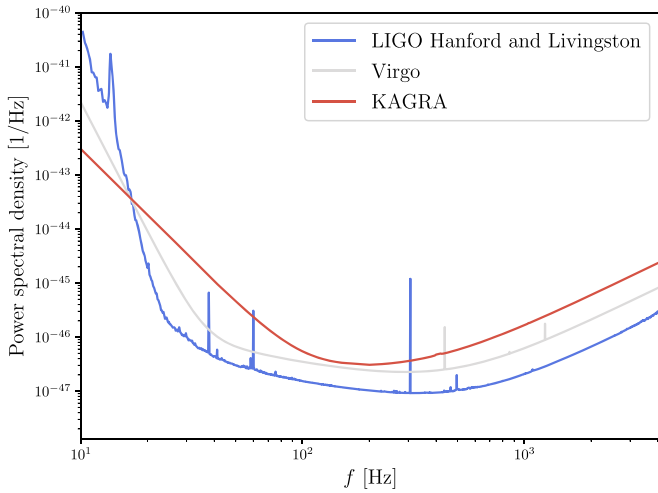
**Table 1**  
Number of BNS Mergers Occurring and Detected in Gravitational Waves for Each Rate Considered in This Work during One Calendar Year of O4

Rate	Mergers	Found	1 day	1 week
Pessimistic	21	$3^{+3}_{-2}$	$0^{+0}_{-0}$	$0^{+1}_{-0}$
Realistic	105	$16^{+6}_{-5}$	$1^{+1}_{-1}$	$6^{+7}_{-5}$
Optimistic	210	$33^{+7}_{-7}$	$4^{+4}_{-3}$	$26^{+19}_{-11}$

**Note.** The second column indicates the total number of mergers, while the third column gives the median and 90% symmetric credible interval on the number of systems detected in gravitational waves obtained by averaging over the 101 different realizations of BNS merger combinations. Also included are the number of unique pairs of events that are found in gravitational waves within one day and one week of each other.

volume. We consider three different merger rates consistent with those presented in the LIGO Scientific Collaboration et al. (2021): pessimistic, realistic, and optimistic rates of 100, 500, and 1000  $\text{Gpc}^{-3} \text{yr}^{-1}$ , respectively. This results in 21, 105, and 210 sources per year out to  $d_L = 400$  Mpc (assuming a merger rate unevolving with redshift; see Table 1). We generate 625 unique BNS merger events following the distributions specified above. For each of the three rates considered, we draw 101 different realizations of the corresponding total yearly number of events from these 625 systems. This allows us to marginalize over many different realizations of the “universe” described by each merger rate and more realistically account for uncertainties.





**Figure 2.** Predicted power spectral densities during O4 of the four interferometers included in this study. These curves correspond to BNS ranges of 160–190 Mpc for the LIGO detectors, 90–120 Mpc for Virgo, and 80 Mpc for KAGRA (Abbott et al. 2020a).

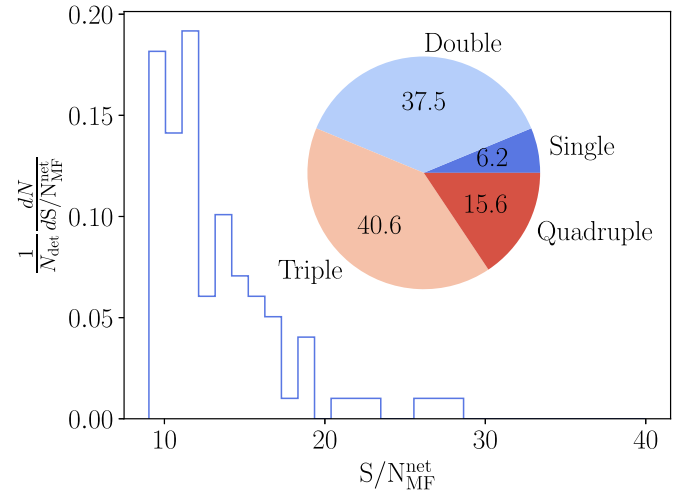
## 2.2. Detection in Gravitational Waves

For each simulated system, we randomly assign a detector configuration assuming each of the four interferometers are independently operating with a duty cycle of 0.7 (Abbott et al. 2020a). Each BNS is added to 296 s of simulated Gaussian noise sampled at 8192 Hz colored by the expected O4 power spectral density (PSD, shown in Figure 2) for each of the detectors that are “observing” during that event (Abbott et al. 2020b), starting at a frequency of 17 Hz.

### 2.2.1. PyCBC Live

Low-latency searches for gravitational waves from compact binary mergers rely on the matched filtering technique (e.g., Cutler et al. 1993; Allen et al. 2012), where a template bank of waveforms is used to filter the data to search for candidate events. In order to more accurately account for the uncertainties associated with the matched-filter detection of gravitational-wave events, we use the PyCBC Live (Nitz et al. 2018; Dal Canton et al. 2021) pipeline—which is one of the pipelines used to search for gravitational-wave events in LIGO and Virgo data in real time—to identify candidates in our simulated population. The discreteness of the template bank is the dominant source of uncertainty on the source parameter estimates obtained from matched filtering. It is constructed using a stochastic placement algorithm (Babak 2008; Harry et al. 2009; Privitera et al. 2014) so that the maximum loss in the signal-to-noise ratio ( $S/N_{\text{GW}}$ ) calculated from a starting frequency of 20 Hz between adjacent templates is 3%. The templates cover detector-frame component masses of between 1 and  $3 M_{\odot}$  and aligned spins out to  $\chi = 0.05$ .

For convenience, the search is configured to run on data from all four detectors in our O4-like network. It calculates the PSD for each detector in real time using the simulated data for each event. A trigger is generated if the criterion  $S/N_{\text{GW}} \geq 4.5$  is met in at least two of the four detectors. In post-processing, we then calculate the network matched-filter  $S/N_{\text{GW}}$  using the trigger information returned by PyCBC Live by taking the square root of the quadrature sum of the  $S/N_{\text{GW}}$  in each detector that was actually “observing” at the time of the event in question.



**Figure 3.** Distribution of the network matched-filter  $S/N_{\text{GW}}$  recovered by PyCBC Live for the 96 total found events. The inset shows the fraction of the total number of found events detected with each number of interferometers.

Because we only simulate data in short segments around the times containing a BNS merger rather than a continuous one-year data stream, the false alarm rate (FAR) cannot be meaningfully calculated by PyCBC Live. Rather than using an FAR threshold to indicate detection, we instead set a threshold on the network matched-filter  $S/N_{\text{GW}}$  recovered by the search for each event of  $S/N_{\text{GW}} > 9$  in order for the source to count as “found.” Of the 625 independent BNS mergers, 96 are found by PyCBC Live. By averaging over the 101 different realizations of BNS merger combinations for each rate as described in the previous section, we obtain the median and 90% symmetric credible interval on the number of found events shown in Table 1. The fraction of the total number of events detected with each number of interferometers is shown in Figure 3, along with the recovered  $S/N_{\text{GW}}$  distribution. The farthest event is found at a distance of 372.2 Mpc and the nearest at 59.0 Mpc.

### 2.2.2. BAYESTAR Map Generation

We localize each event identified by PyCBC Live (i.e., with network  $S/N_{\text{GW}} > 9$ ) using the BAYESTAR rapid localization algorithm (Singer & Price 2016). During O3, skymaps produced using BAYESTAR were released in low-latency for each GW event that passed the public-alert threshold.

BAYESTAR takes as inputs the estimated masses of the neutron stars, the coalescence time, and the  $S/N_{\text{GW}}$  time series from each detector as calculated by PyCBC Live. From this time series, BAYESTAR extracts the timing, relative phases, and amplitudes from each detector. BAYESTAR then constructs a three-dimensional sky localization posterior distribution, with a probability for each pixel on the two-dimensional sky and an additional distance component that is approximated by a Gaussian along each line of sight.

In O3 and in this work, BAYESTAR sky localizations for each event were computed in  $\mathcal{O}(10 \text{ s})$ . Further efforts to optimize BAYESTAR have led to improvements resulting in runtimes of  $\mathcal{O}(1 \text{ s})$  (Magee et al. 2021).

### 2.2.3. Parameter Estimation

After the initial low-latency skymap using the BAYESTAR algorithm is sent out, further source characterization is

performed using full parameter estimation, whereby posterior probability distributions for the binary parameters are obtained using stochastic sampling methods. While BAYESTAR fixes the intrinsic parameter values to those corresponding to the maximum-likelihood template returned by the search pipeline, parameter estimation enables marginalization over the uncertainty in the intrinsic parameters and the extrinsic parameters that are not necessary for the skymap. Accounting for this uncertainty and allowing for correlations between parameters as done using full parameter estimation should lead to skymaps that have systematically smaller sky areas and include the true location of the source at smaller confidence intervals (Singer et al. 2014). We present a comparison of the skymaps from the two localization algorithms in Section 3.1.

The likelihood of observing gravitational-wave data  $d$  given binary parameters  $\theta$  is given by (Veitch & Vecchio 2010; Romano & Cornish 2017; Ashton et al. 2019):

$$\mathcal{L}(d|\theta) \propto \exp\left(-\sum_k \frac{2|d_k - h_k(\theta)|^2}{TS_k}\right), \quad (2)$$

where  $h(\theta)$  represents the gravitational waveform for the BNS signal with parameters  $\theta$ ,  $T$  is the duration of the analyzed segment,  $S_k$  is the PSD, and  $k$  indicates the frequency dependence of the data, waveform, and PSD. The posterior probability distribution for the binary parameters characterizing each systems is then given by Bayes' theorem:

$$p(\theta|d) \propto \mathcal{L}(d|\theta)\pi(\theta), \quad (3)$$

where  $\pi(\theta)$  represents the prior probability distribution assumed for  $\theta$ .

We simulate the progression of skymaps that could occur during a real observing run by performing parameter estimation using the `bilby` software (Ashton et al. 2019; Romero-Shaw et al. 2020) for all of the events that were found by PyCBC Live. We use the PyMultiNest nested sampler (Feroz & Hobson 2008; Feroz et al. 2009, 2019; Buchner et al. 2014) to generate samples from the posterior probability distribution for the source parameters, where the likelihood is analytically marginalized over the distance to the source and the phase at coalescence (Veitch et al. 2015; Singer & Price 2016; Thrane & Talbot 2019). We make the assumption that the neutron stars are point masses with tidal deformability  $\Lambda = 0$ , which is unlikely to affect the inference of the sky location of the source. This assumption enables the use of the reduced order quadrature implementation (Smith et al. 2016) of the IMRPhenomPv2 waveform (Hannam et al. 2014; Husa et al. 2016; Khan et al. 2016), significantly reducing the computational cost of the parameter estimation. The likelihood uses the true PSD used to color the Gaussian noise in each detector. In order to keep the computational cost low, we do not marginalize over uncertainty in the detector calibration.

Because the search pipelines recover the chirp mass of the source extremely well (Biscoveanu et al. 2019), we use a uniform prior on the detector-frame chirp mass with a width of  $0.2 M_\odot$  centered on the true value. The prior on the mass ratio,  $q \equiv m_2/m_1$ , is uniform over the interval  $[0.125, 1]$ . We restrict the analysis to aligned spins only, and the spin prior is the same as that used for drawing the simulated sources, shown in Figure 1. The luminosity distance prior is uniform in the source frame between 10 and 500 Mpc, and the prior on the coalescence time in the geocentric frame is uniform over a

width of 0.2 s centered on the true coalescence time. We use standard priors on all other parameters (see, e.g., Romero-Shaw et al. 2020).

With these parameter estimation settings, the sampling stage takes  $63^{+61}_{-32}$  mins across all of the recovered events. This depends most sensitively on the  $S/N_{\text{GW}}$  of the signal, since louder signals take longer to analyze due to the way the nested sampler explores the prior volume. We note that there are additional timing overheads for generating the weights for calculating the reduced order quadrature likelihood, reconstructing the posterior for the analytically marginalized distance parameter, and generating a skymap of the appropriate format that can be used by observers, although these additional steps generally take less time than the sampling. This timing is not necessarily indicative of what will occur in O4. There are a number of alternative techniques being explored that can be used to further accelerate parameter estimation, particularly for low-mass sources, like focused reduced order quadrature (Morisaki & Raymond 2020), relative binning (Zackay et al. 2018; Finstad & Brown 2020), parallelization (Pankow et al. 2015; Talbot et al. 2019; Wysocki et al. 2019; Smith et al. 2020), prior restrictions (You et al. 2021), and machine-learning-based approaches (Gabbard et al. 2022; Green & Gair 2020; Green et al. 2020; Williams et al. 2021).

## 2.3. Electromagnetic Follow-up

### 2.3.1. Lightcurve Modeling

In order to capture some of the current uncertainty in kilonova modeling, we use two different prescriptions to predict the lightcurves of kilonovae from our simulated BNS mergers.

**Bulla model:** The first is a grid of kilonova models generated using the Monte Carlo radiative transfer code POSSIS (Bulla 2019). The grids assume a two-component ejecta comprised of mass ejected on a dynamical timescale and mass ejected in the form of a wind from the debris disk post merger. The dynamical ejecta is further divided into a lanthanide-rich equatorial component and a lanthanide-free polar component. The lightcurves are calculated using four parameters: the total dynamical ejecta mass ( $M_{\text{ej}}^{\text{dyn}}$ ), the total wind ejecta mass ( $M_{\text{ej}}^{\text{wind}}$ ), the opening angle of the lanthanide-rich component ( $\Phi$ ) and the observer viewing angle ( $\theta_{\text{obs}} = \theta_{\text{JN}}$  if  $\theta_{\text{JN}} < 90^\circ$  and  $\theta_{\text{obs}} = 180^\circ - \theta_{\text{JN}}$  if  $\theta_{\text{JN}} > 90^\circ$ ). A larger value of the opening angle  $\Phi$  increases the amount of lanthanide-rich material, making the kilonova redder. Further details about the BNS grids are presented in Dietrich et al. (2020).

**Kasen model:** The second model uses the kilonova simulations presented in Kasen et al. (2017). While these simulations only include the effects of a single, spherically symmetric ejecta component with radial density given by a broken power law, they feature a more rigorous treatment of the microphysics determining the radiation transport in the system and allow for a range of compositions. The model parameters are the total ejecta mass, which we calculate as  $M_{\text{ej}} = M_{\text{ej}}^{\text{dyn}} + M_{\text{ej}}^{\text{wind}}$ , the expansion velocity of the ejecta,  $v_{\text{ej}}$ , and the mass fraction of lanthanides,  $X_{\text{lan}}$ . A higher  $M_{\text{ej}}$  leads to lightcurves that peak at brighter magnitudes, while higher  $v_{\text{ej}}$  leads to faster-fading lightcurves. More neutron-rich ejecta with a higher lanthanide fraction,  $X_{\text{lan}} \gtrsim 10^{-2}$ , produce a redder and longer-lived kilonova than ejecta composed primarily of light  $r$ -process material,  $10^{-6} \lesssim X_{\text{lan}} \lesssim 10^{-2}$  (Kasen et al. 2017).

To estimate the ejecta masses and velocity from the neutron star masses, we use the prescription outlined in Stachie et al. (2021). First, we assume an AP4 equation of state (Akmal & Pandharipande 1997; Akmal et al. 1998) for the merging neutron stars to calculate their radii and compactness, although recent works have developed (Stachie et al. 2021) and employed (Abbott et al. 2021a) methods for marginalizing over the uncertainty in the equation of state. We then calculate the dynamical ejecta mass using the fitting formula from Coughlin et al. (2019a):

$$\log_{10} M_{\text{ej}}^{\text{dyn}} = \left[ a \frac{(1 - 2C_i) m_i}{C_i} + b m_2 \left( \frac{m_1}{m_2} \right)^n + \frac{d}{2} \right] + [1 \leftrightarrow 2], \quad (4)$$

where  $m_i$  and  $C_i$  are the masses and the compactnesses of the two neutron stars, respectively, and  $a = -0.0719$ ,  $b = 0.2116$ ,  $d = -2.42$ , and  $n = -2.905$ . The ejecta velocity is similarly calculated using a fitting formula from Coughlin et al. (2019a):

$$v_{\text{ej}} = \left[ a'' \frac{m_1}{m_2} (1 + c'' C_i) + \frac{b''}{2} \right] + [1 \leftrightarrow 2], \quad (5)$$

where the coefficients are given by  $a'' = -0.3090$ ,  $b'' = 0.657$ , and  $c'' = -1.879$ .

We estimate the wind ejecta mass as a fraction of the disk mass  $M_{\text{ej}}^{\text{wind}} = \zeta M_{\text{ej}}^{\text{disk}}$  and set  $\zeta = 0.15$  (Dietrich et al. 2020). We calculate the disk mass using the fitting formula

$$\log_{10} M_{\text{ej}}^{\text{disk}} = \max \left[ -3, a' \left( 1 + b' \tanh \left( \frac{c' - (m_1 + m_2)/M_{\text{thresh}}}{d'} \right) \right) \right], \quad (6)$$

where  $M_{\text{thresh}}$  is the minimum total mass that results in a prompt collapse post merger and is calculated as in Bauswein et al. (2013);  $a'$ ,  $b'$ ,  $c'$ , and  $d'$  are calculated as in Dietrich et al. (2020).

We interpolate between the parameters of standard grids for both the Bulla and Kasen models using the Python package *gwemlightcurves* (Coughlin et al. 2018a, 2019a; Dietrich et al. 2020) to predict the  $J$ -band and  $r$ -band lightcurves of the kilonovae. The left panel of Figure 4 shows the lightcurves for an example system. The differences in peak magnitude and decay time for the lightcurves calculated with the Kasen and Bulla models demonstrate the uncertainties currently underlying kilonova modeling. We present results for both models to conservatively account for this uncertainty in our predictions. Additionally, we note that the ejecta masses calculated from the fitting formulae are approximate and depend strongly on the equation of state, fitting parameters, and assumptions of the numerical relativistic models that they are based on. It is thus possible that the ejecta masses are not entirely representative of realistic kilonovae. We investigate the effect of varying ejecta masses in Section 4.1.

### 2.3.2. WINTER Simulated Observing

For each found gravitational-wave event, we use the corresponding skymap and simulated time of the event to create a realistic observing schedule with the *gwemopt* package (Ghosh et al. 2016; Coughlin et al. 2018b, 2019b).

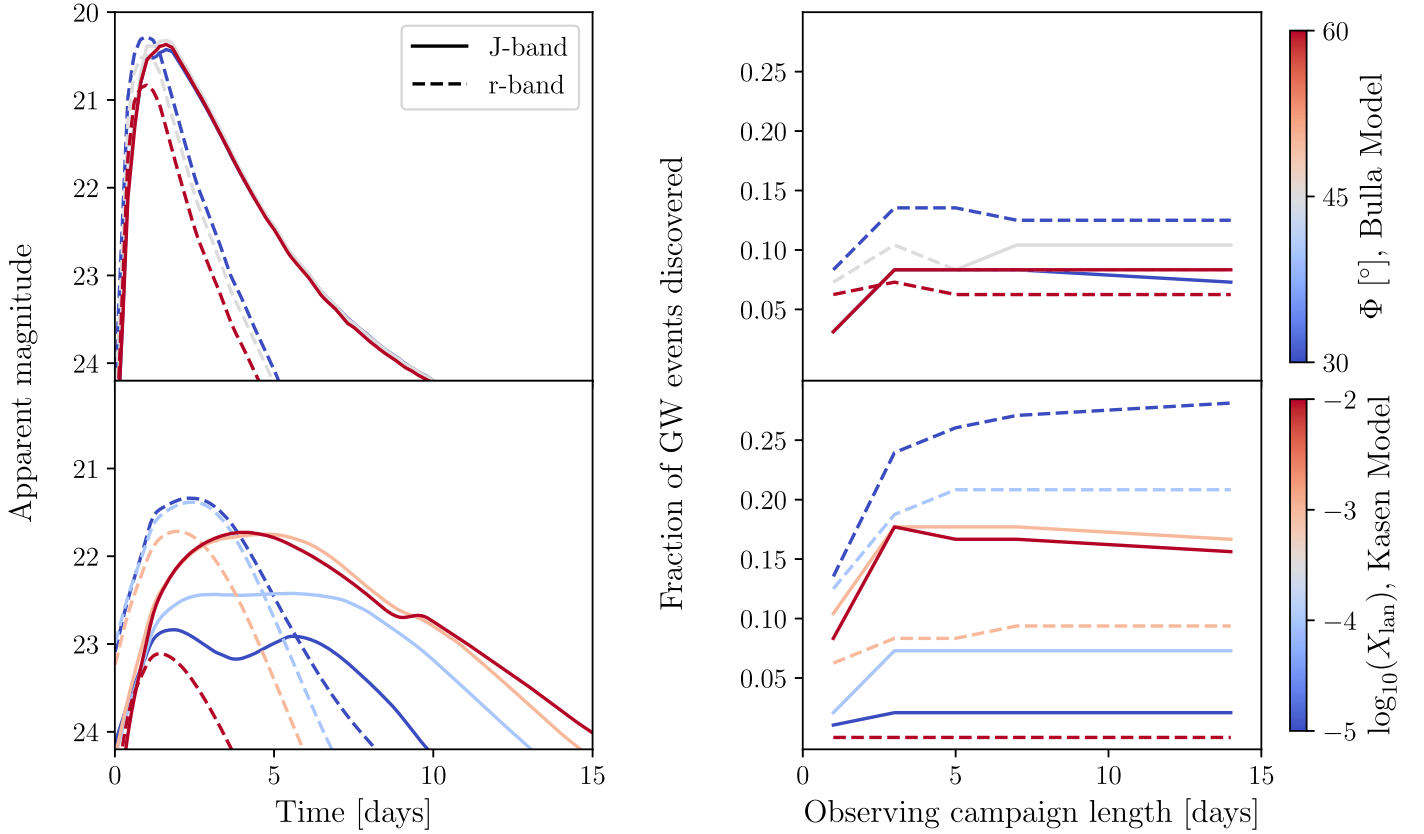
The package takes gravitational-wave probability maps as inputs, such as the *BAYESTAR* and *bilby* skymaps, subdivides the sky into tiles sized to the telescope field of view, and generates an optimized observing schedule. During O3, a network of telescopes including the Zwicky Transient Facility (Bellm et al. 2018) created GW follow-up schedules with *gwemopt* (e.g., Coughlin et al. 2019c; Kasliwal et al. 2019), and we expect to use the package during nominal WINTER operations in O4.

We run a set of observing simulations allowing one, three, five, seven, and 14 nights of dedicated telescope time searching for the kilonova from each event. For the scope of this study, we limit observations to the  $J$  band to match WINTER's planned  $J$ -band reference images and all-sky survey and only study one gravitational-wave event at a time, even though multiple events may occur during the same night or same week (see Table 1). Each WINTER observation lasts 450 s to match the  $J_{AB} = 21$  mag reference images, split across five dithers, with the time for dithering at approximately one second per dither represented as an overhead time in the simulation. Time to slew between each field is calculated based on the telescope and dome slew rates measured at WINTER's host telescope at Palomar Observatory. WINTER's InGaAs sensors read out continuously during each exposure, leading to no overhead time due to sensor readout (Malonis et al. 2020). Combining overhead and exposure times, for an eight hour night WINTER covers  $\sim 63 \text{ deg}^2$  to  $J_{AB} = 21$  mag. The WINTER data processing pipeline will subtract new science images from prebuilt reference images (constructed well before the GW alert) and detect candidate kilonovae in near real time.

We follow up each skymap with a ranked search strategy, in which we subdivide the sky into a fixed grid of telescope pointings. The center of each tile corresponds to a WINTER reference image for easy image subtraction during data processing. We prioritize the tiles based on the skymap-generated probability. Finally, we create a greedy observing schedule, where the highest probability tiles are observed first, as described in Coughlin et al. (2018b). We prioritize scheduling at least two observations of each field, spread out over the length of an observing campaign to study the lightcurve evolving over time. In this simplified simulation, one observing schedule is created and executed without modification for each follow-up campaign. The simulation schedules multiple visits for each field, but fields are not preferentially revisited. During real O4 follow-up observing, WINTER will iteratively compare observations to prebuilt reference images and revisit any fields containing promising kilonova candidates (see Section 4.3 for details).

Next, we check the observing schedules for a successful observation of the kilonova, which can be defined in multiple ways. For the sake of clarity in this study, we define an event as *localized* if WINTER takes at least one image of the kilonova's true location on the sky. An event may not be localized if the skymap is too large for WINTER to search it efficiently or if the event is not overhead at the given time of year. Additionally, poor weather can hinder follow-up observations; however, weather simulations are outside of the scope of this study.

Even if an event is localized, the kilonova must be sufficiently bright at the time of imaging to be detected. The event qualifies as a *discovery* if it is localized, imaged at least twice, and observed to a signal-to-noise ratio ( $S/N_{\text{EM}} \geq 5$ ). We



**Figure 4.** Variation due to lightcurve models. Left: example lightcurves in the  $J$  (solid lines) and  $r$  (dashed lines) filters for the Bulla model (top) and the Kasen model (bottom) with different lanthanide fractions as shown by the top and bottom color bars, respectively. The example BNS system has masses  $m_1 = 1.36 M_\odot$ ,  $m_2 = 1.22 M_\odot$  and is at a distance  $d_L = 162.97$  Mpc. Right: the fraction of events discovered in the  $J$  band by WINTER and an equally sensitive  $r$ -band telescope, as described in Section 2.3.2, plotted against the number of nights allocated to a kilonova search, marginalized over all BNS merger event rates. Unique lightcurves are modeled for each event from the same set of Bulla (top) and Kasen (bottom) models.

employ the lightcurve models described in Section 2.3.1 to calculate the magnitude of the event at the time of each observation and scale the magnitude based on the airmass of the observation. We follow Equation (1) as described in Section 3.1 of Frostig et al. (2020) to calculate the  $S/N_{EM}$  for each event based on historic Palomar data and predicted instrument noises. For comparison, we repeat this exercise for WINTER observing in  $J$  band and for a fictitious optical telescope observing in the Sloan Digital Sky Survey  $r$  filter with equivalent sensitivity, field of view, exposure times, and overhead times as the WINTER  $J$ -band observations.

For the scope of this study, we use two observations to  $S/N_{EM} \geq 5$  as a simplified proxy for true kilonova discovery, as two observations are the minimum number required to distinguish a kilonova from asteroids or other transients. In reality, to confirm a new kilonova candidate, more than two observations may be required, and promising candidates will be selected based on their color evolution, how quickly the lightcurve fades, and if the event is associated with a probable host galaxy. Discovery will then be confirmed with further photometric and spectroscopic follow-up. See Section 4.3 for further discussion of electromagnetic follow-up observations.

### 3. Results

#### 3.1. Skymap Comparisons

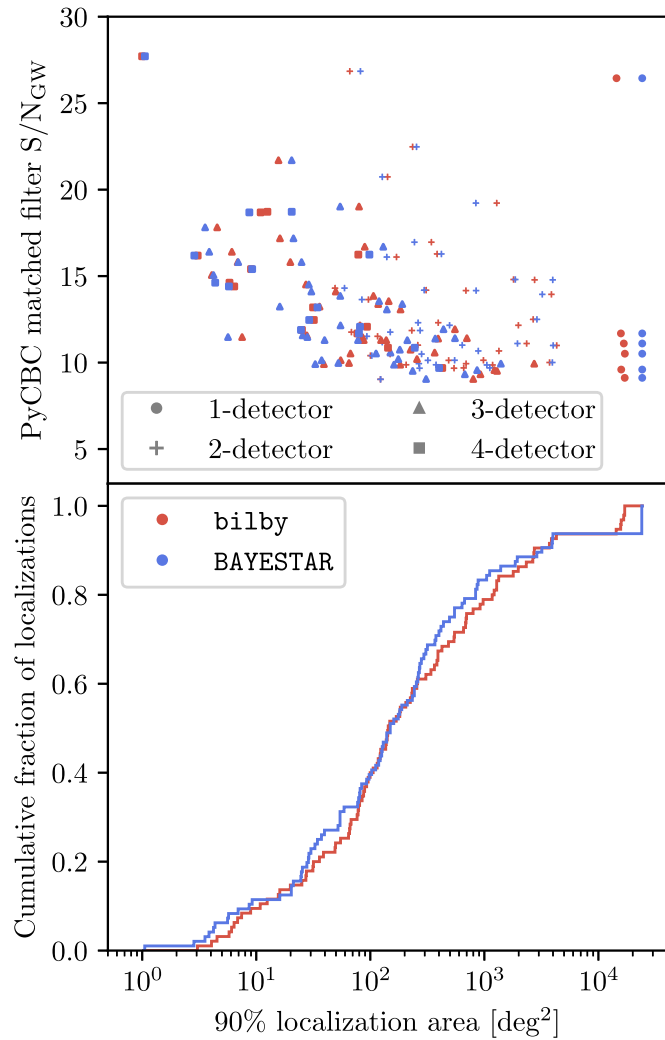
A total of 96 events were found by PyCBC Live out of the 625 total independent simulated mergers, as shown in Table 1.

Each event was localized by both BAYESTAR and bilby, as described in Section 2. We present a comparison of the two algorithms. This is especially relevant for electromagnetic follow-up purposes, since in O3, BAYESTAR skymaps were distributed hours or days before the more comprehensive bilby or LALInference (Veitch et al. 2015) skymap. If this delay persists in O4 and if there are significant discrepancies in the two localizations, electromagnetic observers might switch partway through the night, or might even prefer to wait for the bilby skymap before beginning observations.

The distribution of 90% credible localization areas for all bilby and BAYESTAR skymaps is shown in Figure 5. Events that are recovered by PyCBC Live with larger values of matched-filter  $S/N_{GW}$  tend to have smaller localization areas. Some events have network  $S/N_{GW} > 20$  but localizations larger than  $100 \text{ deg}^2$ ; these are typically one- or two-detector events where the low number of detectors results in poor constraints on timing and phase. The cumulative distributions of 90% localization areas from each algorithm are similar, with 24% of bilby and 27% of BAYESTAR localizations falling under  $50 \text{ deg}^2$ . The BAYESTAR distribution lies slightly to the left (i.e., to smaller areas) of the bilby distribution.

Figure 6 shows a comparison between the BAYESTAR and bilby skymaps for two particular events that are generally indicative of the performance observed in the larger sample. In the left skymap, which has a 90% localization area of order  $\mathcal{O}(1000) \text{ deg}^2$  from a two-detector event at a distance





**Figure 5.** Top: 90% credible localization areas of bilby and BAYESTAR skymaps for each event vs. its PyCBC Live matched-filter  $S/N_{\text{GW}}$ . Different markers represent the number of GW detectors that identified the event. Bottom: cumulative distribution of 90% credible localization areas for skymaps generated by both algorithms.

$d_L = 141.1$  Mpc, the BAYESTAR and bilby skymaps have some overlap, but with probability concentrated in different areas of the sky. In the right skymap, which stems from a four-detector event at a distance  $d_L = 113.8$  Mpc, the localizations are small and in very good agreement.

A useful metric for the accuracy of localizations of simulated events is the searched area, which is the amount of sky area that is covered by integrating in order of decreasing probability from the highest probability pixel until reaching the true location of the source. The left panel of Figure 7 shows the distributions of searched areas for bilby and BAYESTAR. The two algorithms perform very similarly.

Another convenient statistic for comparing probability distributions is the Jensen–Shannon (JS) divergence. The JS divergence measures how similar two distributions are and ranges from 0 bit for identical distributions to 1 bit for completely divergent distributions. JS divergence values greater than 0.002 are considered to be statistically significant (Romero-Shaw et al. 2020). The right panel of Figure 7 shows the intersection of the two 90% credible areas normalized by the bilby 90% area versus the JS divergence

between the two skymap posterior distributions for each event. These distributions describe the probability of the event being in each pixel of the two-dimensional skymap. For skymaps that have almost complete 90% area overlap, the JS divergence is very small and thus the maps share significant information content, and vice versa. However, there are also events for which the normalized intersections are close to unity, but have relatively large JS divergences. This is due to two types of events: events that have similar 90% localization areas, but with different probability distributions for the pixels within that area; and events for which the intersection of the bilby 90% area is small compared to the BAYESTAR 90% area, so the intersection divided by the bilby area is almost unity.

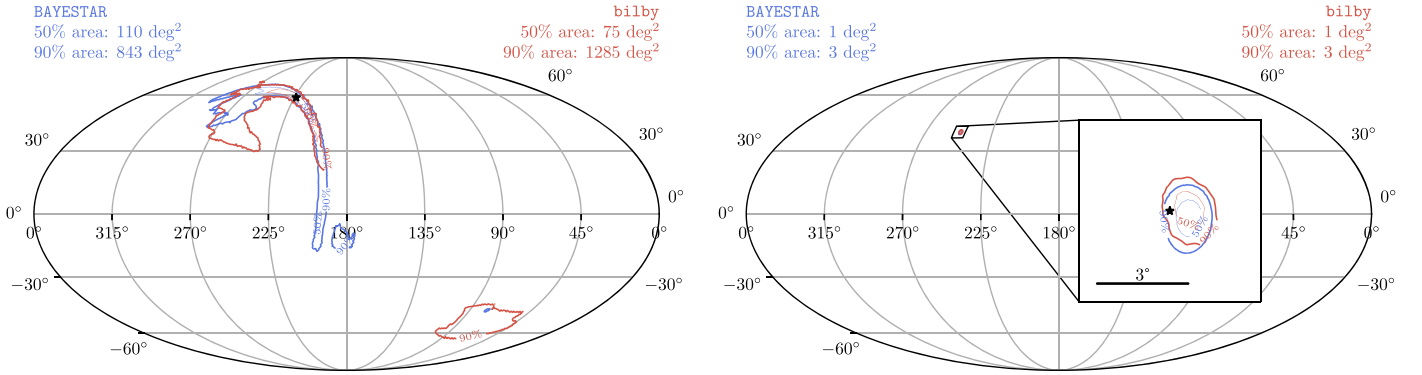
Previous studies have found that localizations from full parameter estimation pipelines such as LALInference, when compared to those from BAYESTAR, have systematically smaller sky areas and include the true location of the source at smaller confidence intervals (see, e.g., Figure 3 of Singer et al. 2014). In this study we find that the discrepancy between the two algorithms is significantly reduced, since our BAYESTAR skymaps use data from all online detectors, while those in Singer et al. (2014) use data only from detectors that register an  $S/N_{\text{GW}}$  above the detection threshold. Furthermore, since matched-filter searches recover the true chirp mass of BNS systems with extremely high accuracy, to within  $\sim \mathcal{O}(10^{-4}) M_{\odot}$  (Biscoveanu et al. 2019), not marginalizing over the uncertainty in the mass parameters should lead to less significant biases in the low-latency skymap compared to higher-mass sources. We expect the difference between the skymaps obtained with the two algorithms to be more significant for NSBH sources, which will be explored in future work.

### 3.2. Results from WINTER Simulated Observing

Figure 8 shows the results of searching the bilby and BAYESTAR maps with WINTER, varying the length of the observing campaign from one through 14 nights of telescope time, given an optimistic BNS merger rate. The simulation produces separate observing schedules based on the amount of time allowed to follow up the event. In agreement with Section 3.1, WINTER localizes the BNS merger events at similar rates for both the bilby and BAYESTAR skymaps. The number of events localized with WINTER steadily increases given more nights of observing time, with  $\sim 2$  times more events localized with a 14-night search strategy than a one-night search strategy.

Varying the number of nights allowed searching for the kilonova changes the ordering and prioritization of each observing schedule. For example, given a one-night observing campaign, 100% of the images of the kilonova are taken before the  $J$ -band peak of the lightcurve, regardless of model. For a five-night campaign, 80% (86%) of the localized events have their first images taken before the peak and 30% (27%) have two images taken before the peak for the Kasen (Bulla) lightcurve model. Increasing observing time allows WINTER to cover a greater portion of the skymap, but risks tiling the high-probability area of the skymap less efficiently or observing the kilonova later in its evolution when it may have already faded. Therefore, discovery of new kilonovae (defined above as observing the event at least twice to  $S/N_{\text{EM}} \geq 5$ ) does not increase steadily with time, but levels off as the kilonova eventually fades beyond detection. In the right panel of Figure 4, we compare discovery of new kilonovae in the





**Figure 6.** Comparison sky maps showing the 50% and 90% credible regions obtained using *bilby* and BAYESTAR for two different events. The true source location is marked with the black star. The event on the left was observed with the Hanford and KAGRA interferometers at a distance of 141 Mpc, although it was not confidently detected with KAGRA since the optimal  $S/N_{\text{GW}}$  in that detector is 2.6. The one on the right was detected with Hanford, Livingston, Virgo, and KAGRA at a distance of 114 Mpc. In the sky map to the right, the BAYESTAR and *bilby* localizations almost completely overlap.

WINTER *J* band and for an equally sensitive *r*-band telescope given the number of nights each telescope is allowed to search for the kilonovae.

Given the Kasen lightcurve models, discovery in both *J*-band and *r*-band peaks around three nights of tiling and levels off given more nights of observing, with the exception of lanthanide-poor ( $X_{\text{lan}} = 10^{-5}$ ) kilonova discovery in the *r* band continuing to increase with 14 nights of observing (bottom right panel of Figure 4). In the *J* band with three nights of observing, WINTER discovers 8.5 times more lanthanide-rich ( $X_{\text{lan}} = 10^{-2}$ ) events than lanthanide-poor events ( $X_{\text{lan}} = 10^{-5}$ ). By contrast, an equally sensitive *r*-band telescope discovers zero lanthanide-rich events and up to 28% of lanthanide-poor events.

Despite differences in the lightcurve models, *J*-band and *r*-band discovery also levels off after three nights of observing for the Bulla models (top right panel of Figure 4). Varying the opening angle of the lanthanide-rich component changes the resultant lightcurves less than varying the lanthanide fraction in the Kasen models. At three nights of observing, a large opening angle of the lanthanide-rich component ( $\Phi = 60^\circ$ ) decreases *r*-band discovery by 46% and does not change *J*-band discovery.

Table 2 displays the results of the end-to-end simulation studying how many kilonovae WINTER will discover during one year of follow-up observations, including the median and 90% symmetric credible interval on the number of events as described in Section 2.2.1. Some subset of the GW triggers will not be observable by a telescope at Palomar Observatory at the given trigger time, either because the event is too far south or too near the Sun. For a given year of GW triggers, on average  $\sim 70\%$  of the events are visible above  $20^\circ$  altitude for WINTER at some point during the night following the event trigger time. These events are denoted as EM accessible in the table. For a portion of those events, WINTER observes the correct patch of sky and localizes the events, with some events missed due to large localization areas or events with true locations outside of the 90% localization areas. Localized events only qualify as a discovery ( $S/N_{\text{EM}} \geq 5$  in at least two observations) if the event is sufficiently bright, which can vary based on the model grid used to simulate the event. WINTER discoveries range from as low as zero new kilonovae per year with a pessimistic BNS merger rate to as high as 10 new kilonovae discovered per year to 90% confidence with an optimistic BNS merger rate. Given a realistic BNS merger rate, WINTER discovers up to five new kilonovae per year to 90% confidence.

## 4. Discussion

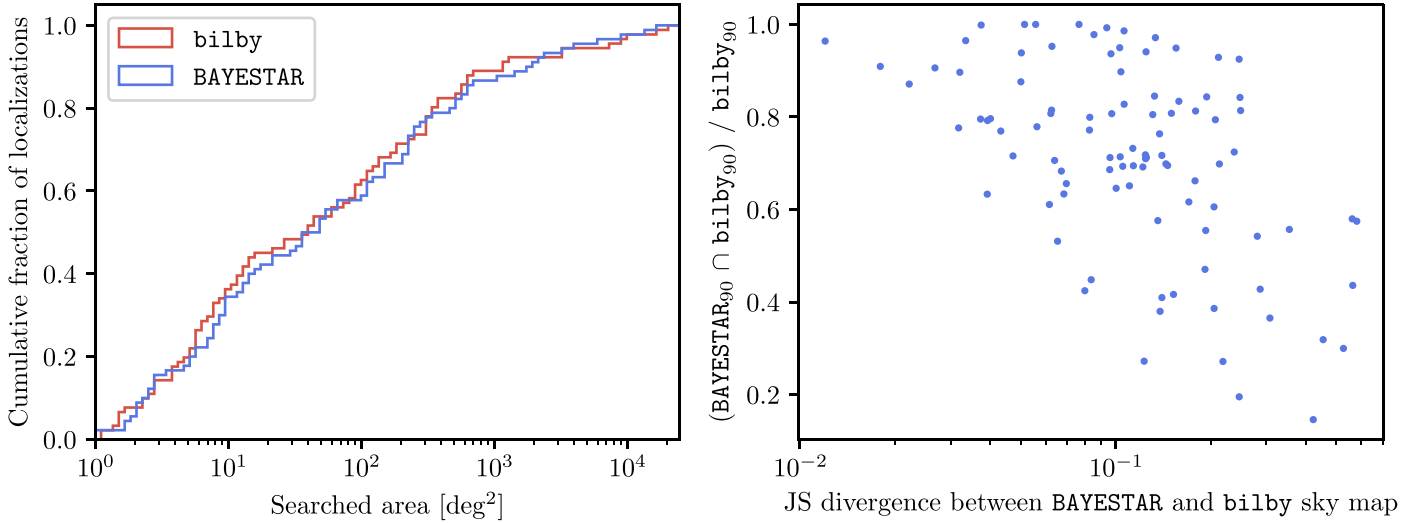
### 4.1. Advantages of Infrared Follow-up

In Section 3, we demonstrate that a 1 deg<sup>2</sup> *J*-band survey like WINTER can detect up to 10 kilonovae during O4. Here, we examine the advantages of a *J*-band search over a similar optical search. We use a grid of realistic kilonova lightcurves calculated with the Bulla model to identify the parameter space where an infrared search outperforms an optical search. To generate the grid, we do not use the fitting formulae from Section 2.3.1, as they may not be entirely representative of the underlying physical population of kilonovae. Instead, we calculate lightcurves for a wide range of possible ejecta masses derived from numerical relativistic simulations:  $M_{\text{ej}}^{\text{dyn}} = [0.001, 0.01] M_\odot$  and  $M_{\text{ej}}^{\text{wind}} = [0.01, 0.13] M_\odot$  (Andreoni et al. 2020). We set the dynamical ejecta opening angles in the range  $\Phi = [30^\circ, 60^\circ]$  and calculate lightcurves for viewing angles sampled uniformly in  $\cos(\theta_{\text{obs}})$ .

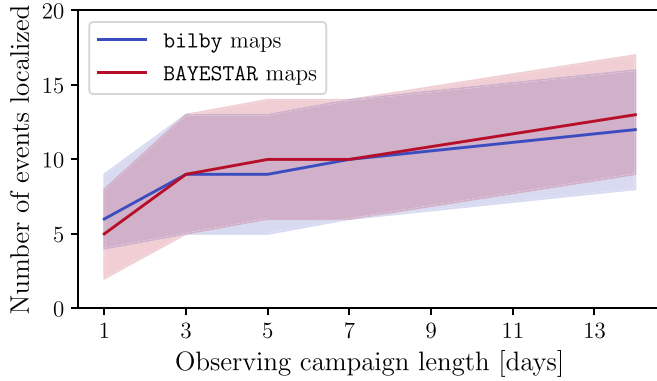
#### 4.1.1. Red Kilonovae Are Brighter at Infrared Wavelengths

Kilonovae with a larger opening angle of the dynamical ejecta will have more lanthanide-rich material and hence will be brighter at redder wavelengths. We quantify this in Figure 9, which shows the maximum distance out to which a kilonova can be detected in the *r* and *J* bands by a telescope with a limiting depth of 21 mag. We distinguish between blue ( $\Phi = 30^\circ$ ) and red ( $\Phi = 60^\circ$ ) kilonovae, and an on-axis ( $\theta_{\text{obs}} < 60^\circ$ ) and off-axis ( $\theta_{\text{obs}} > 60^\circ$ ) kilonova. We use the maximum-likelihood estimates of the ejecta masses of GW170817 as a benchmark ( $M_{\text{ej}}^{\text{dyn}} \sim 0.005 M_\odot$ ,  $M_{\text{ej}}^{\text{wind}} \sim 0.05 M_\odot$ ; Dietrich et al. 2020) and indicate the distances out to which a kilonova with these masses can be detected. Finally, we plot contours corresponding to the distances of 15 events from one realization of our uniformly distributed, realistic-rate simulation that were detected in GWs and followed up with WINTER (see Table 1). Twelve of these events are on-axis and three are off-axis.

For a blue kilonova ( $\Phi = 30^\circ$ ), an infrared search does not offer a significant advantage over an optical search. A blue on-axis (off-axis) kilonova with GW170817-like ejecta would be detectable out to  $\approx 300$  (230) Mpc in both *r* and *J* bands. This is because the peak values of *r*- and *J*-band lightcurves for a blue kilonova are not significantly different (see top left panel of Figure 4). Off-axis kilonovae are generally fainter than on-axis



**Figure 7.** Left: cumulative distributions of searched area for skymaps generated using *bilby* and BAYESTAR. Right: the intersection of the BAYESTAR 90% and *bilby* 90% credible areas, normalized by the *bilby* 90% credible area vs. Jensen–Shannon divergences between *bilby* and BAYESTAR skymaps for each event.



**Figure 8.** The number of events localized by WINTER for an optimistic BNS merger rate given the length of the observing campaign for following up each event with the *bilby* and BAYESTAR skymaps. The solid lines represent the median number of events localized and the shaded regions show the 90% symmetric credible interval.

ones, which explains the reduced sensitivity (300 Mpc versus 220 Mpc) for off-axis kilonovae. All 12 on-axis events from our simulation are detectable in both  $r$  and  $J$  bands if they are blue. We note that the detectable region of the ejecta mass phase space (i.e., the region to the right of each contour in Figure 9) is slightly larger for  $r$  band than for the  $J$  band. Ten of these events have a detectable region that includes GW170817 ejecta masses in both  $r$  and  $J$  bands. All three off-axis events from our simulation are detectable in both  $r$  and  $J$  bands. Of these three, two are closer than 150 Mpc, and are detectable for the entire range of ejecta masses in both  $r$  and  $J$  bands.

However, an infrared search performs significantly better than an optical search for red kilonovae. Both on- and off-axis red kilonovae are detectable out to larger distances in the infrared than in the optical. A red on-axis (off-axis) kilonova with GW170817-like ejecta is detectable to 284 (246) Mpc in the infrared but only to 217 (174) Mpc in the optical. If the 12 on-axis kilonovae from our simulations are red, only 10 are detectable in the  $r$  band while all 12 are detectable in the  $J$  band. GW170817 ejecta masses lie in the detectable region for only six kilonovae in the  $r$  band but for 10 kilonovae in  $J$  band. If the three off-axis simulated events are red, only two are detectable in the  $r$  band, while all three can be detected in the  $J$

band. Finally, we note that if a particular kilonova is not detected in WINTER observations, Figure 9 can be used to place constraints on the ejecta masses and opening angles associated with it.

#### 4.1.2. All Kilonovae Are Longer Lived in the Infrared

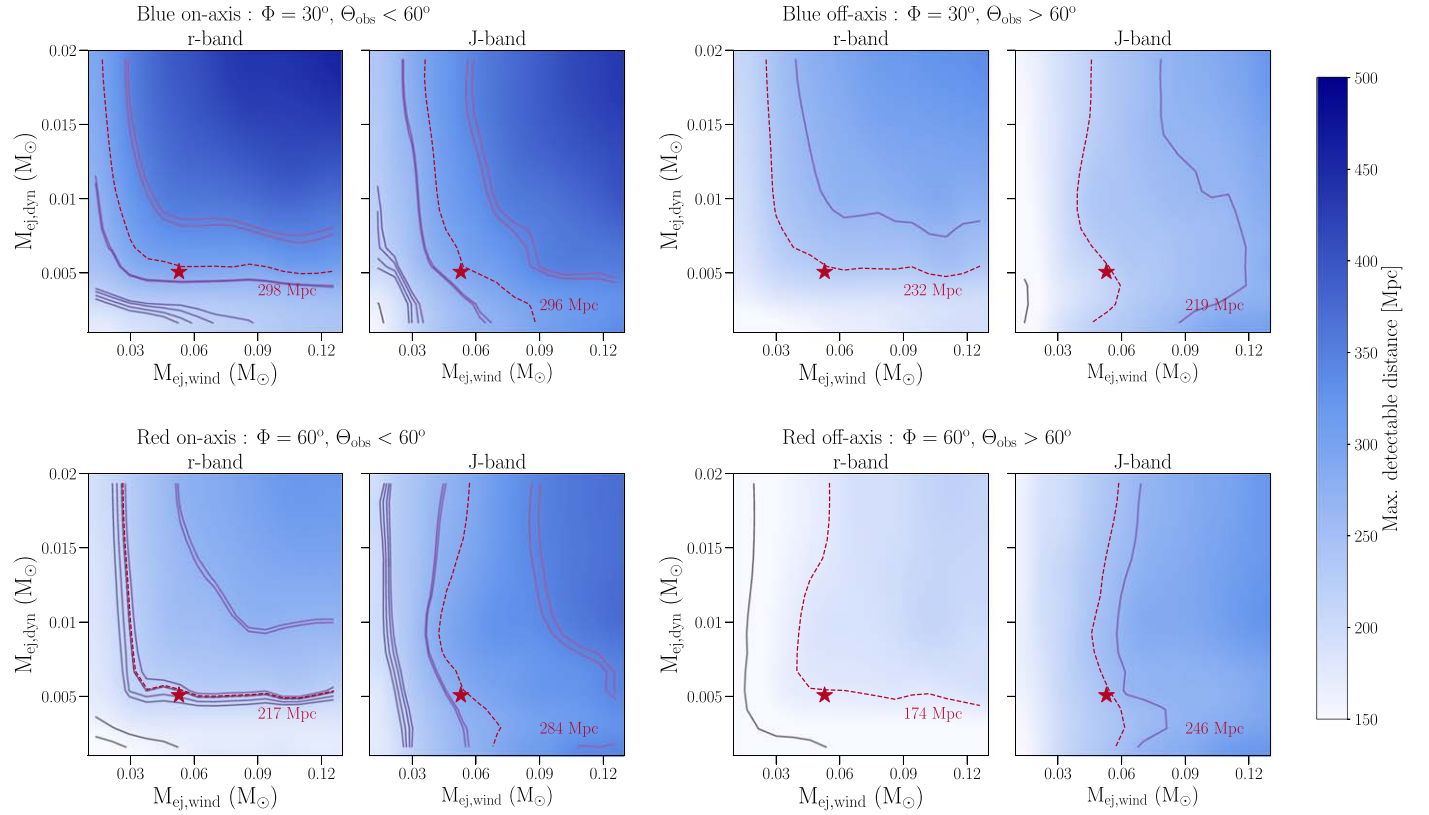
A second advantage of infrared searches is that kilonova emission is longer lived in the infrared than in the optical. We quantify this in Figure 10, which shows the fraction of kilonovae from our entire model grid that can be detected in the  $r$  and  $J$  bands by an  $m_{\text{lim}} = 21$  telescope as a function of distance and number of days since the merger. We plot contours corresponding to detection fractions ( $f_{\text{det}}$ ) of 0.1, 0.5, and 0.9.

Figure 10 clearly shows that kilonovae can be detected for much longer in the  $J$  band compared to  $r$  band. For example, at 200 Mpc more than 10% (i.e.,  $f_{\text{det}} > 0.1$ ) of kilonovae are detectable in the  $r$  band for a maximum duration of three days. However, in the  $J$  band the same events can be detected for almost six days. A WINTER-like telescope with a  $1 \text{ deg}^2$  field of view can thus search localization regions of  $\approx 450 \text{ deg}^2$  in the  $J$  band for kilonovae at 200 Mpc, but is limited to only  $\approx 200 \text{ deg}^2$  in the  $r$  band. At a distance of 100 (300) Mpc, these values change to  $\approx 350$  (120)  $\text{deg}^2$  for  $r$  band and 750 (220)  $\text{deg}^2$  for  $J$  band.

In Figure 10, we also plot the distances and areas enclosing 90% localization probabilities from the skymaps of the 15 events from one realization of our realistic BNS rate simulation. For an  $r$ -band search, seven of the 15 events have 90% areas lying in the  $f_{\text{det}} > 0.1$  region. In the  $J$  band, eight events lie in this region.

#### 4.2. WINTER GW Follow-up Strategy

WINTER is a dedicated instrument for follow-up of slowly fading infrared kilonovae and can spend weeks searching for a single event. However, even though longer searches cover larger fractions of the skymaps, they risk covering the high-probability areas less efficiently or observing the kilonova once it has already faded. Based on the results from simulating WINTER observations with the Kasen and Bulla models,



**Figure 9.** The maximum distance out to which a kilonova can be detected by an  $m_{\text{lim}} = 21$  survey for a given combination of dynamical and wind ejecta masses, based on the Bulla models. We plot the distances separately for  $J$  and  $r$  bands. We further distinguish the kilonova models as blue and on-axis (top left), blue and off-axis (top right), red and on-axis (bottom left) and red and off-axis (bottom right). The ejecta masses for GW170817 are plotted as a red star. The red dashed line shows the contour at the distance at which a kilonova with GW170817-like ejecta masses is detectable in each case (distance indicated in red). We also plot the contours at the distances of 15 events followed up by WINTER from one realization of our realistic-rates simulation. Twelve of these events are on-axis and three are off-axis. Two off-axis events have distances  $< 150$  Mpc and lie off the plots. It is evident that for the same set of ejecta masses, the  $J$  band can detect kilonovae out to larger distances than  $r$  band if the kilonova is red.

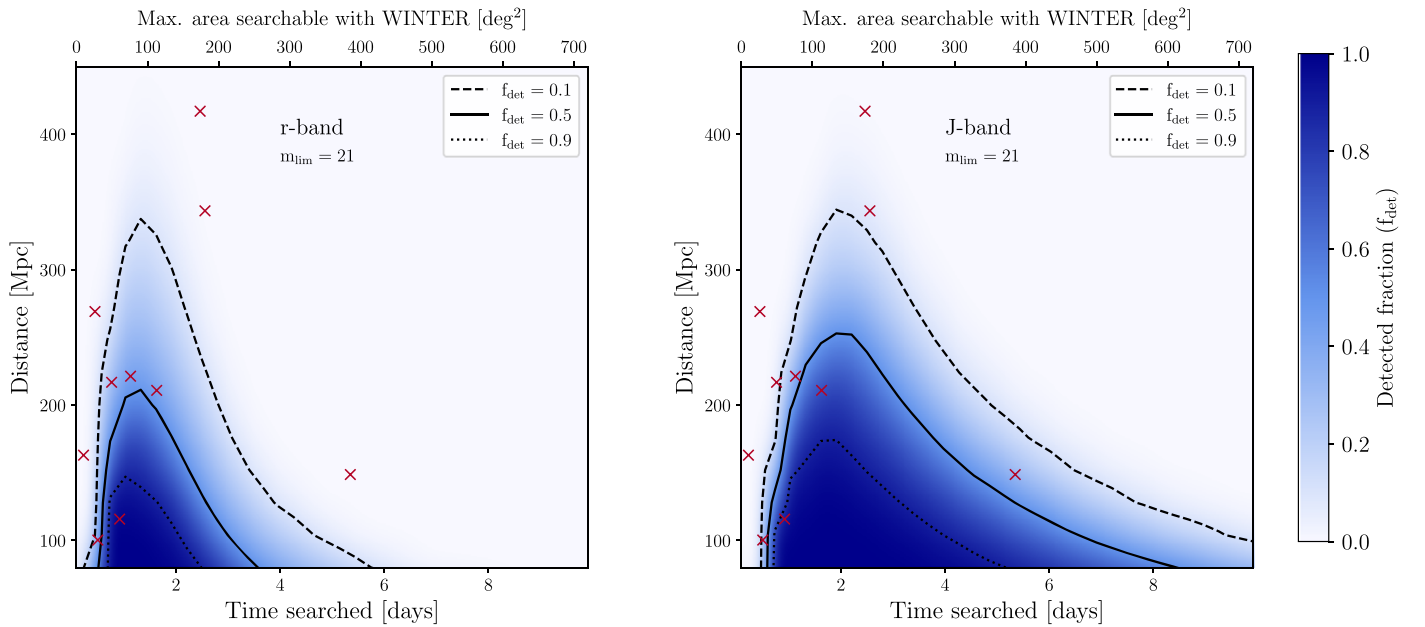
**Table 2**

Number of Gravitational-wave Triggers Leading to Various Categories of WINTER Observations Given Pessimistic, Realistic, and Optimistic Event Rates

Rate	GW Triggers	EM Accessible	Localized	Discovered			
				Bulla		Kasen	
	Events	Events	Events	$\Phi$ ( $^\circ$ )	Events	$X_{\text{lan}}$	Events
Pessimistic	$3^{+3}_{-2}$	$2^{+2}_{-2}$	$1^{+1}_{-1}$	30	$0^{+1}_{-0}$	$10^{-2}$	$0^{+2}_{-0}$
				45	$0^{+1}_{-0}$	$10^{-3}$	$0^{+2}_{-0}$
				60	$0^{+1}_{-0}$	$10^{-4}$	$0^{+1}_{-0}$
						$10^{-5}$	$0^{+0}_{-0}$
Realistic	$16^{+6}_{-5}$	$11^{+5}_{-3}$	$5^{+3}_{-3}$	30	$1^{+2}_{-1}$	$10^{-2}$	$2^{+3}_{-2}$
				45	$1^{+2}_{-1}$	$10^{-3}$	$3^{+2}_{-2}$
				60	$1^{+2}_{-1}$	$10^{-4}$	$1^{+2}_{-1}$
						$10^{-5}$	$0^{+1}_{-0}$
Optimistic	$33^{+7}_{-7}$	$23^{+5}_{-7}$	$10^{+4}_{-4}$	30	$3^{+1}_{-2}$	$10^{-2}$	$6^{+3}_{-4}$
				45	$3^{+2}_{-2}$	$10^{-3}$	$6^{+4}_{-3}$
				60	$3^{+2}_{-2}$	$10^{-4}$	$2^{+2}_{-2}$
						$10^{-5}$	$1^{+1}_{-1}$

**Note.** An event is accessible if it is overhead at Palomar Observatory at the given time of year, it is localized if the telescope takes an image of the kilonova's location on the sky, and the kilonova is discovered if detected at least twice to  $S/N_{\text{EM}} \geq 5$  in  $J$  band. All simulations in this table use the ranked search follow-up strategy with five nights of searching the BAYESTAR skymaps.





**Figure 10.** The fraction of kilonovae from our entire model grid that can be detected in the  $r$  and  $J$  bands by an  $m_{\text{lim}} = 21$  telescope, as a function of distance and number of days since the merger. The contours corresponding to detection fractions ( $f_{\text{det}}$ ) of 0.1, 0.5, and 0.9 are plotted as dashed, solid, and dotted lines, respectively. Red crosses mark the distances and areas enclosing 90% localization probability of the events from our realistic-rate simulation (of the 16 events, only 10 are shown, as the remaining six lie outside the bounds of the axes). It is clear that kilonovae can be detected for much longer in the  $J$  band compared to  $r$  band. The right panel can also be used to select GW triggers that are worth following with WINTER. We will only follow events that have median distance estimates and localization areas such that the detection fraction is at least 10%. We note that the detection fraction drops at very early search times ( $t < 0.5$  day), as the kilonova is still brightening in our models at these times. However, we will observe events with localization areas that can be tiled within 0.5 day, observing them repeatedly until the kilonova becomes bright enough to be detected.

WINTER will dedicate up to seven nights of searching for each event. WINTER will observe compelling transients until they fade, and if there are no candidate kilonovae in the data, will stop searching after seven nights.

Furthermore, WINTER will not follow up all BNS GW alerts. Figure 10 provides a prescription to select which GW triggers are worth following up with WINTER during O4 based on information that is available at the time of the trigger. We will only follow events that have median distance estimates and localization areas such that the chance of detecting a kilonova is at least 10% (i.e., lying within the  $f_{\text{det}} > 0.1$  region of Figure 10). With a depth of  $J_{AB} = 21$  (matching the WINTER reference images with a single exposure time  $t_{\text{exp}} = 450$  s), this means we can follow up events with distances of up to 350 Mpc if the time to tile the localization area is less than two days. We can follow up nearby events for much longer, with tiling times of six days for 200 Mpc. With WINTER’s 1 deg<sup>2</sup> field of view, this corresponds to  $\approx 150$  deg<sup>2</sup> at 350 Mpc and 300 deg<sup>2</sup> at 250 Mpc.

If the localization areas are larger and the events are nearer, we will reduce our exposure times to tile the localizations faster and increase the chances of detecting a kilonova. Figure 11 shows the  $f_{\text{det}} = 0.1$  contours for searches with depths of 21 mag ( $t_{\text{exp}} = 450$  s), 20.5 mag ( $t_{\text{exp}} = 180$  s), and 19.5 mag ( $t_{\text{exp}} = 40$  s) as a function of the localization area that can be searched with WINTER. With WINTER, we will follow up all events with localization areas  $< 300$  deg<sup>2</sup> down to a depth of 21 mag, matching WINTER’s  $J$ -band reference images. For events with localization areas larger than 300 deg<sup>2</sup> and distances  $< 250$  Mpc, we will reduce exposure time to 180 s (i.e., to a depth of 20.5 mag). With 180 s exposures, we can tile areas as large as 1000 deg<sup>2</sup> while maintaining  $f_{\text{det}} = 0.1$ . For events with localization areas larger than 1000 deg<sup>2</sup>, we will follow only

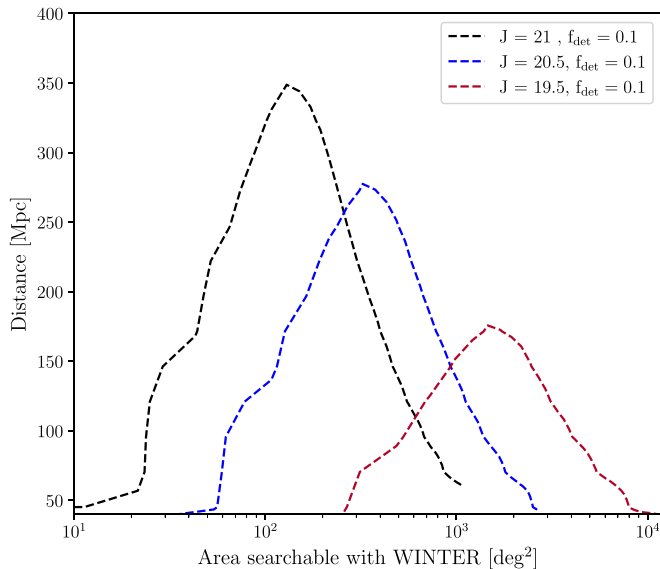
those events that are closer than 150 Mpc to a depth of  $J_{AB} = 19.5$  mag. If there are multiple gravitational-wave triggers of interest on the same night (as discussed in Table 1), we will prioritize events that have a higher  $f_{\text{det}}$  and are easier to observe with WINTER.

Finally, we have assumed that the areas mentioned above enclose all of the BNS localization probability. If the full skymap areas are larger than the limits mentioned above, we will cover only those events where the area enclosing 50% of the localization probability can be tiled with  $f_{\text{det}} > 0.2$ . We also note that in Figure 10, the detection fraction  $f_{\text{det}}$  drops at very early search times ( $t < 0.5$  day), as the kilonova is still brightening in our models at these times. However, we will observe events with localization areas that can be tiled within 0.5 day, observing them repeatedly until the kilonova becomes bright enough to be detected. Similar detectability constraints were derived by Chase et al. (2021; their Figure 17) using the Los Alamos National Laboratory (LANL, Wollaeger et al. 2021) grid of kilonova models. The constraints presented in Figure 10 are broadly consistent with their constraints, with minor variations attributable to the differences in the underlying Bulla and LANL models (see, for example, the differences in the analyses of Anand et al. 2021 and Thakur et al. 2020; Dichiara et al. 2021).

The methods described above can be used by other surveys to select GW triggers for follow-up during O4. We include a python notebook with the code to reproduce Figures 9–11 in the Zenodo repository at Frostig et al. (2021).

#### 4.3. The Realities of Electromagnetic Follow-up Observing

The methods outlined in Section 2.3.2 represent a simplified approach to follow-up observations, where one observing strategy is decided at the outset and followed unchangingly



**Figure 11.** Contours marking the regions for which the kilonova detection fraction is at least 0.1 ( $f_{\text{det}} > 0.1$ ), as a function of the localization area. We plot separate contours for searches with limiting magnitude of 21, 20.5, and 19.5 mag (black, blue, and red, respectively). With WINTER, we will follow up all events with localization areas  $< 450 \text{ deg}^2$  down to a depth of 21 mag. If the localization area is larger than  $450 \text{ deg}^2$ , we will follow only those events that have median distance estimates  $< 200 \text{ Mpc}$  down to a depth of 20.5 mag. If the localization area is larger than  $1000 \text{ deg}^2$ , we will follow only those events with median distances  $< 150 \text{ Mpc}$  down to a depth of 19.5 mag.

throughout the campaign. In reality, follow-up observing is an iterative process where new decisions are made nightly or even multiple times per night. Instead of solely following one search strategy as shown in the above simulations, we will prioritize repeat observations of compelling transients found in the follow-up data from WINTER and other telescopes. In simulations, WINTER observes at the zenith to a limiting magnitude of  $J_{AB} = 21$  in 450 s,  $J_{AB} = 21.8$  in 30 minutes, and  $J_{AB} = 22.7$  in 3 hr, with many repeat observations allowing for more significant constraints on kilonova detection.

Furthermore, if another telescope discovers a new kilonova candidate, infrared localizations, not just discoveries, add unique data to kilonova science. A nondetection in WINTER of an optical kilonova constrains both the dynamical and wind ejecta masses (Chase et al. 2021). Additionally, faint detections ( $S/N_{\text{EM}} < 5$ ) or single images of the event in the near-infrared contribute to models of chemical evolution, ejecta mass, and wind speeds (Kasen et al. 2013; Barnes et al. 2016; Coughlin et al. 2020c).

However, with no confirmed discoveries from other telescopes, there can be thousands of candidate transient events to sort through in a week of near-infrared survey data. There are many tools available for classifying transients in survey data with machine-learning algorithms becoming a standard tool in the field, particularly for wide-field optical surveys, such as ZTF and the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; Goldstein et al. 2015; Mahabal et al. 2019; Gómez et al. 2020; Sooknunan et al. 2020; Turpin et al. 2020; Förster et al. 2021).

Carefully planning the color and cadence of WINTER follow-up observations can assist these transient classification techniques in narrowing down the number of candidate events. For example, the characteristic  $\sim 1$  week fading time distinguishes a kilonova lightcurve from longer-lasting

supernovae or short-lived asteroids (Cowperthwaite & Berger 2015). Additionally, observations in at least two filters assist in studying the reddening of the transient over time, a distinctive feature of kilonovae seen in GW170817 (Arcavi et al. 2017; Pian et al. 2017; Smartt et al. 2017). For optical telescopes, the  $i$  band provides the reddest images, and models predict kilonova discovery is maximized with the  $g-i$  filter pair for LSST (Andreoni et al. 2019) and a  $g$ ,  $r$ , and  $i$  filter cycle for ZTF (Almualla et al. 2021). In the simulation described in Section 2.3.2, we only observe in the  $J$  band to leverage WINTER’s  $J$ -band reference images and all-sky survey. In practice, we aim to conduct most of the search in  $J$  band but also follow up all interesting candidate events in the  $Y$  band to study the  $Y-J$  color evolution. We will also study the  $g-J$  color pair, either with  $g$ -band images from ZTF follow-up or with the optical camera on WINTER’s companion port. Time permitting, WINTER and its counterpart optical camera will also observe candidate events in the  $u$ ,  $r$ ,  $i$ , and short- $H$  filters.

## 5. Conclusion

The BNS merger GW170817 brought about a new field of multimessenger astronomy, but despite extensive follow-up campaigns in O3, we have not observed a second multimessenger kilonova. In this study, we show infrared observations are a promising avenue for kilonova discovery, particularly for lanthanide-rich “red” kilonovae, as these are detectable to larger distances in the infrared than at optical wavelengths. We predict that infrared follow-up of GW triggers with WINTER could discover up to 10 new kilonovae per year during O4. Furthermore, by employing more targeted follow-up strategies than those we have simulated, we can achieve a deeper sensitivity on a subset of interesting targets, therefore enhancing our ability to confirm new discoveries.










Moreover, we limit this study to BNS mergers and leave the study of NSBH kilonovae to a future work. Infrared follow-up of NSBH kilonovae is especially promising, as they are brighter in the infrared compared to BNS kilonovae (Fernández et al. 2017; Zhu et al. 2020; Anand et al. 2021), and observing both event types has the potential to increase the number of kilonovae discovered each year. Even just one new electromagnetic observation of a kilonova in O4 will double the number of known multimessenger kilonovae, helping to answer ongoing questions in the study of the Hubble tension, the neutron star equation of state, and  $r$ -process nucleosynthesis.

The authors thank Michael Coughlin for his support integrating WINTER into `gwemopt` and for help with `gwemlightcurves`. WINTER’s construction is made possible by the National Science Foundation under MRI grant No. AST-1828470. We also acknowledge significant support for WINTER from the California Institute of Technology, the Caltech Optical Observatories, the Bruno Rossi Fund of the MIT Kavli Institute for Astrophysics and Space Research, and the MIT Department of Physics and School of Science. S.B., G.M., H.-Y.C., E.K., and S.V. acknowledge support of the National Science Foundation and the LIGO Laboratory. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under cooperative agreement PHY-0757058. S.V. is supported by the NSF through award PHY-2045740. S.B. is also supported by the NSF Graduate Research Fellowship under grant No.

DGE-1122374. G.M. is supported by the NSF through award PHY-1764464. M.M.K. acknowledges generous support from the David and Lucille Packard Foundation. The authors are grateful for computational resources provided by the LIGO Lab and supported by NSF Grants PHY-0757058 and PHY-0823459. This paper carries LIGO document number LIGO-P2100340.

*Software:* astropy (Astropy Collaboration et al. 2013), gwemopt (Coughlin et al. 2018b), gwemlightcurves (Coughlin et al. 2018a, 2019a; Dietrich et al. 2020), matplotlib (Hunter 2007), numpy (Harris et al. 2020), ligo.skymap (Singer & Price 2016; Singer et al. 2016a, 2016b), pandas (McKinney et al. 2010), bilby (Ashton et al. 2019; Romero-Shaw et al. 2020), PyCBC Live (Nitz et al. 2018; Dal Canton et al. 2021), PyMultinest (Feroz & Hobson 2008; Feroz et al. 2009, 2019; Buchner et al. 2014).

### ORCID iDs

Danielle Frostig  <https://orcid.org/0000-0002-7197-9004>  
 Sylvia Biscoveanu  <https://orcid.org/0000-0001-7616-7366>  
 Geoffrey Mo  <https://orcid.org/0000-0001-6331-112X>  
 Viraj Karambelkar  <https://orcid.org/0000-0003-2758-159X>  
 Tito Dal Canton  <https://orcid.org/0000-0001-5078-9044>  
 Hsin-Yu Chen  <https://orcid.org/0000-0001-5403-3762>  
 Mansi Kasliwal  <https://orcid.org/0000-0002-5619-4938>  
 Robert A. Simcoe  <https://orcid.org/0000-0003-3769-9559>  
 Salvatore Vitale  <https://orcid.org/0000-0003-2700-0767>

### References

- Aasi, J., Abbott, B. P., LIGO Scientific Collaboration, et al. 2015, *CQGra*, **32**, 074001
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, *ApJL*, **848**, L12
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017d, *Natur*, **551**, 85
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, *LRR*, **21**, 3
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020a, *LRR*, **23**, 3
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020b, Noise Curves Used for Simulations in the Update of the Observing Scenarios Paper, LIGO-T2000012-v2 NSF, <https://dcc.ligo.org/LIGO-T2000012/public>
- Abbott, B. P., LIGO Scientific Collaboration, Virgo Collaboration, et al. 2017a, *PhRvL*, **119**, 161101
- Abbott, B. P., LIGO Scientific Collaboration, Virgo Collaboration, et al. 2017c, *PhRvL*, **119**, 161101
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021b, *PhRvX*, **11**, 021053
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021a, *ApJL*, **915**, L5
- Abbott, R., LIGO Scientific Collaboration, Virgo Collaboration, et al. 2021b, arXiv:2108.01045
- Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, *CQGra*, **32**, 024001
- Ackley, K., Amati, L., Barbieri, C., et al. 2020, *A&A*, **643**, A113
- Akmal, A., & Pandharipande, V. R. 1997, *PhRvC*, **56**, 2261
- Akmal, A., Pandharipande, V. R., & Ravenhall, D. G. 1998, *PhRvC*, **58**, 1804
- Akutsu, T., Ando, M., Araki, S., et al. 2018, *PTEP*, **2018**, 013F01
- Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. 2012, *PhRvD*, **85**, 122006
- Almualla, M., Anand, S., Coughlin, M. W., et al. 2021, *MNRAS*, **504**, 2822
- Alpar, M. A., Cheng, A. F., Ruderman, M. A., & Shaham, J. 1982, *Natur*, **300**, 728
- Anand, S., Coughlin, M., Kisiwal, M. M., et al. 2021, *NatAs*, **5**, 46
- Andreoni, I., Anand, S., Bianco, F. B., et al. 2019, *PASP*, **131**, 068004
- Andreoni, I., Kool, E. C., Sagués Carracedo, A., et al. 2020, *ApJ*, **904**, 155
- Antier, S., Agayeva, S., Almualla, M., et al. 2020, *MNRAS*, **497**, 5518
- Arcavi, I., Hosseinzadeh, G., Howell, A. D., et al. 2017, *Natur*, **551**, 64
- Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, *ApJS*, **241**, 27
- Aso, Y., Michimura, Y., Somiya, K., et al. 2013, *PhRvD*, **88**, 043007
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. P., et al. 2013, *A&A*, **558**, A33
- Babak, S. 2008, *CQGra*, **25**, 195011
- Barnes, J., & Kasen, D. 2013, *ApJ*, **775**, 18
- Barnes, J., Kasen, D., Wu, M.-R., & Martínez-Pinedo, G. 2016, *ApJ*, **829**, 110
- Bauswein, A., Baumgarte, T. W., & Janka, H. T. 2013, *PhRvL*, **111**, 131101
- Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2018, *PASP*, **131**, 018002
- Becerra, R. L., Dichiaro, S., Watson, A. M., et al. 2021, *MNRAS*, **507**, 1401
- Biscoveanu, S., Vitale, S., & Haster, C.-J. 2019, *ApJL*, **884**, L32
- Breschi, M., Perego, A., Bernuzzi, S., et al. 2021, *MNRAS*, **505**, 1661
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, **564**, A125
- Bulla, M. 2019, *MNRAS*, **489**, 5037
- Burgay, M., D'Amico, N., Possenti, A., et al. 2003, *Natur*, **426**, 531
- Catelan, M., Minniti, D., Lucas, P. W., et al. 2011, in *RR Lyrae Stars, Metal-Poor Stars, and the Galaxy*, ed. A. McWilliam, Vol. 5 (Pasadena, CA: The Observatories of the Carnegie Institution of Washington)
- Chase, E. A., O'Connor, B., Fryer, C. L., et al. 2021, arXiv:2105.12268
- Chen, H.-Y., Fishbach, M., & Holz, D. E. 2018, *Natur*, **562**, 545
- Coughlin, M. W., Tao, D., Chan, M. L., et al. 2018b, *MNRAS*, **478**, 692
- Coughlin, M. W., Ahumada, T., Cenko, B. S., et al. 2019c, *PASP*, **131**, 048001
- Coughlin, M. W., Antier, S., Corre, D., et al. 2019b, *MNRAS*, **489**, 5775
- Coughlin, M. W., Dietrich, T., Antier, S., et al. 2020b, *MNRAS*, **497**, 1181
- Coughlin, M. W., Dietrich, T., Antier, S., et al. 2020c, *MNRAS*, **492**, 863
- Coughlin, M. W., Dietrich, T., Doctor, Z., et al. 2018a, *MNRAS*, **480**, 3871
- Coughlin, M. W., Dietrich, T., Heinzel, J., et al. 2020a, *PhRvR*, **2**, 022006
- Coughlin, M. W., Dietrich, T., Margalit, B., & Metzger, B. D. 2019a, *MNRAS*, **489**, L91
- Coulter, D. A., Foley, R. J., Kilpatrick, C. D., et al. 2017, *Sci*, **358**, 1556
- Cowperthwaite, P. S., & Berger, E. 2015, *ApJ*, **814**, 25
- Cutler, C., Apostolatos, T. A., Bildsten, L., et al. 1993, *PhRvL*, **70**, 2984
- Dal Canton, T., Nitz, A. H., Gadre, B., et al. 2021, *ApJ*, **923**, 254
- De, K., Hankins, M. J., Kasliwal, M. M., et al. 2020, *PASP*, **132**, 025001
- Dichiaro, S., Becerra, R. L., Chase, E. A., et al. 2021, *ApJ*, **923**, L32
- Dietrich, T., Coughlin, M. W., Pang, P. T. H., et al. 2020, *Sci*, **370**, 1450
- Dobie, D., Stewart, A., Murphy, T., et al. 2019, *ApJL*, **887**, L13
- Evans, P. A., Cenko, S. B., Kennea, J. A., et al. 2017, *Sci*, **358**, 1565
- Farrow, N., Zhu, X.-J., & Thrane, E. 2019, *ApJ*, **876**, 18
- Fernández, R., Foucart, F., Kasen, D., et al. 2017, *CQGra*, **34**, 154001
- Feroz, F., & Hobson, M. P. 2008, *MNRAS*, **384**, 449
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, **398**, 1601
- Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, *OJAp*, **2**, 10
- Finstad, D., & Brown, D. A. 2020, *ApJL*, **905**, L9
- Fong, W., Laskar, T., Rastinejad, J., et al. 2021, *ApJ*, **906**, 127
- Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., et al. 2021, *AJ*, **161**, 242
- Frostig, D., Baker, J. W., Brown, J., et al. 2020, Proc. SPIE, 11447, 1144767
- Frostig, D., Biscoveanu, S., Mo, G., et al. 2021, Public Release of Data Associated with WINTER/LIGO BNS Kilonova Simulations v1.0, Zenodo, doi:10.5281/zenodo.5507322
- Gabbard, H., Messenger, C., Heng, I. S., Tonolini, F., & Murray-Smith, R. 2022, *NatPh*, **18**, 112
- Ghosh, S., Bloemen, S., Nelemans, G., Groot, P. J., & Price, L. R. 2016, *A&A*, **592**, A82
- Goldstein, A., Veres, P., Burns, E., et al. 2017, *ApJL*, **848**, L14
- Goldstein, D. A., D'Andrea, C. B., Fischer, J. A., et al. 2015, *AJ*, **150**, 165
- Gómez, C., Neira, M., Hernández-Hoyos, M., Arbeláez, P., & Forero-Romero, J. E. 2020, *MNRAS*, **499**, 3130
- Gompertz, B. P., Cutter, R., Steeghs, D., et al. 2020, *MNRAS*, **497**, 726
- Green, S. R., & Gair, J. 2020, arXiv:2008.03312
- Green, S. R., Simpson, C., & Gair, J. 2020, *PhRvD*, **102**, 104057
- Grossman, D., Korobkin, O., Rosswog, S., & Piran, T. 2014, *MNRAS*, **439**, 757
- Haggard, D., Nynka, M., Ruan, J. J., et al. 2017, *ApJL*, **848**, L25
- Hallinan, G., Corsi, A., Mooley, K. P., et al. 2017, *Sci*, **358**, 1579
- Hannam, M., Schmidt, P., Bohé, A., et al. 2014, *PhRvL*, **113**, 151101
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Harry, I. W., Allen, B., & Sathyaprakash, B. S. 2009, *PhRvD*, **80**, 104014
- Heuvel, E. P. J. V. D. 2017, *JApA*, **38**, 45
- Holz, D. E., & Hughes, S. A. 2005, *ApJ*, **629**, 15
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Hu, Y. D., Li, X. Y., Castro-Tirado, A. J., et al. 2021, in VI Workshop on Robotic Autonomous Observatories, ed. A. J. Castro-Tirado, S. B. Pandey, & M. D. Caballero-García (Mexico City: Instituto de Astronomía), **75**
- Husa, S., Khan, S., Hannam, M., et al. 2016, *PhRvD*, **93**, 044006
- Kasen, D., Badnell, N. R., & Barnes, J. 2013, *ApJ*, **774**, 25
- Kasen, D., Metzger, B., Barnes, J., Quataert, E., & Ramirez-Ruiz, E. 2017, *Natur*, **551**, 80
- Kasliwal, M. M., Anand, S., Ahumada, T., et al. 2020, *ApJ*, **905**, 145
- Kasliwal, M. M., Cannella, C., Bagdasaryan, A., et al. 2019, *PASP*, **131**, 038003
- Kasliwal, M. M., Kasen, D., Lau, R. M., et al. 2022, *MNRAS*, **510**, L7
- Khan, S., Husa, S., Hannam, M., et al. 2016, *PhRvD*, **93**, 044007
- Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, *PASP*, **129**, 104502



- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, **379**, 1599
- Levan, A. 2020, *PoS, Asterics2019*, 044
- Li, L.-X., & Paczynski, B. 1998, *ApJL*, **507**, L59
- Lipunov, V. M., Gorbvskoy, E., Kornilov, V. G., et al. 2017, *ApJL*, **850**, L1
- Lorimer, D. R. 2008, *LRR*, **11**, 8
- Lourie, N. P., Baker, J. W., Burruss, R. S., et al. 2020, *Proc. SPIE*, **11447**, 114479K
- Lyne, A. G., Burgay, M., Kramer, M., et al. 2004, *Sci*, **303**, 1153
- Magee, R., Chatterjee, D., Singer, L. P., et al. 2021, *ApJL*, **910**, L21
- Mahabal, A., Rebbapragada, U., Walters, R., et al. 2019, *PASP*, **131**, 038002
- Malonis, A. C., Lourie, N. P., Fűrész, G., et al. 2020, *Proc. SPIE*, **11454**, 114542O
- Margalit, B., & Metzger, B. D. 2017, *ApJL*, **850**, L19
- Margutti, R., Berger, E., Fong, W., et al. 2017, *ApJL*, **848**, L20
- McKinney, W., et al. 2010, in *Proc. of the 9th Python in Science Conference*, ed. S. van der Walt & J. Millman (Austin, TX: SciPy), 51
- Metzger, B. D. 2020, *LRR*, **23**, 1
- Metzger, B. D., Martinez-Pinedo, G., Darbha, S., et al. 2010, *MNRAS*, **406**, 2650
- Morisaki, S., & Raymond, V. 2020, *PhRvD*, **102**, 104020
- Nitz, A. H., Dal Canton, T., Davis, D., & Reyes, S. 2018, *PhRvD*, **98**, 024050
- Oates, S. R., Marshall, F. E., Breeveld, A. A., et al. 2021, *MNRAS*, **507**, 1296
- Page, K. L., Evans, P. A., Tohuvavohu, A., et al. 2020, *MNRAS*, **499**, 3459
- Pankow, C., Brady, P., Ochsner, E., & O’Shaughnessy, R. 2015, *PhRvD*, **92**, 023002
- Paterson, K., Lundquist, M. J., Rastinejad, J. C., et al. 2021, *ApJ*, **912**, 128
- Petrov, P., Singer, L. P., Coughlin, M. W., et al. 2022, *ApJ*, **924**, 54
- Pian, E., D’Avanzo, P., Benetti, S., et al. 2017, *Natur*, **551**, 67
- Privitera, S., Mohapatra, S. R. P., Ajith, P., et al. 2014, *PhRvD*, **89**, 024003
- Radhakrishnan, V., & Srinivasan, G. 1982, *CSci*, **51**, 1096
- Roberts, L. F., Kasen, D., Lee, W. H., & Ramirez-Ruiz, E. 2011, *ApJL*, **736**, L21
- Romano, J. D., & Cornish, N. J. 2017, *LRR*, **20**, 2
- Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, *MNRAS*, **499**, 3295
- Rosswog, S. 2005, *ApJ*, **634**, 1202
- Simcoe, R. A., Fűrész, G., Sullivan, P. W., et al. 2019, *AJ*, **157**, 46
- Singer, L. P., Chen, H.-Y., Holz, D. E., et al. 2016a, *ApJL*, **829**, L15
- Singer, L. P., Chen, H.-Y., Holz, D. E., et al. 2016b, *ApJS*, **226**, 10
- Singer, L. P., & Price, L. R. 2016, *PhRvD*, **93**, 024013
- Singer, L. P., Price, L. R., Farr, B., et al. 2014, *ApJ*, **795**, 105
- Smartt, S. J., Chen, T.-W., Jerkstrand, A., et al. 2017, *Natur*, **551**, 75
- Smith, R., Field, S. E., Blackburn, K., et al. 2016, *PhRvD*, **94**, 044031
- Smith, R. J. E., Ashton, G., Vajpeyi, A., & Talbot, C. 2020, *MNRAS*, **498**, 4492
- Soares-Santos, M., Holz, D. E., Annis, J., et al. 2017, *ApJL*, **848**, L16
- Somiya, K. 2012, *CQGra*, **29**, 124007
- Sooknunan, K., Lochner, M., Bassett, B. A., et al. 2020, *MNRAS*, **502**, 206
- Stachie, C., Coughlin, M. W., Dietrich, T., et al. 2021, *MNRAS*, **505**, 4235
- Talbot, C., Smith, R., Thrane, E., & Poole, G. B. 2019, *PhRvD*, **100**, 043030
- Tanaka, M., & Hotokezaka, K. 2013, *ApJ*, **775**, 113
- Tanvir, N. R., Levan, A. J., Fruchter, A. S., et al. 2013, *Natur*, **500**, 547
- Tauris, T. M., Kramer, M., Freire, P. C. C., et al. 2017, *ApJ*, **846**, 170
- Thakur, A. L., Dichiara, S., Troja, E., et al. 2020, *MNRAS*, **499**, 3868
- The LIGO Scientific CollaborationThe Virgo CollaborationThe KAGRA Scientific Collaboration 2021, arXiv:2111.03634
- Thrane, E., & Talbot, C. 2019, *PASA*, **36**, e036
- Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, *PASP*, **130**, 064505
- Travouillon, T. D., Moore, A. M., Soon, J., et al. 2020, *Proc. SPIE*, **11445**, 114454L
- Troja, E., Piro, L., van Eerten, H., et al. 2017, *Natur*, **551**, 71
- Troja, E., Ryan, G., Piro, L., et al. 2018, *NatCo*, **9**, 4089
- Turpin, D., Ganet, M., Antier, S., et al. 2020, *MNRAS*, **497**, 2641
- Veitch, J., & Vecchio, A. 2010, *PhRvD*, **81**, 062003
- Veitch, J., Raymond, V., Farr, B., et al. 2015, *PhRvD*, **91**, 042003
- Williams, M. J., Veitch, J., & Messenger, C. 2021, *PhRvD*, **103**, 103006
- Wollaeger, R. T., Fryer, C. L., Chase, E. A., et al. 2021, *ApJ*, **918**, 10
- Wysocki, D., O’Shaughnessy, R., Lange, J., & Fang, Y.-L. L. 2019, *PhRvD*, **99**, 084026
- You, Z.-Q., Ashton, G., Zhu, X.-J., Thrane, E., & Zhu, Z.-H. 2021, *MNRAS*, **509**, 3957
- Zackay, B., Dai, L., & Venumadhav, T. 2018, arXiv:1806.08792
- Zhu, J.-P., Wu, S., Yang, Y.-P., et al. 2021a, *ApJ*, **921**, 156
- Zhu, J.-P., Wu, S., Yang, Y.-P., et al. 2021b, *ApJ*, **917**, 24
- Zhu, J.-P., Yang, Y.-P., Liu, L.-D., et al. 2020, *ApJ*, **897**, 20